Ioan Iacob Cristian R. Gheorghiu

Elemente de statistică aplicate în educație fizică și sport

EDITURA UNIVERSITĂȚII „AL. I. CUZA"
IAȘI – 1998

Scanned with OKEN Scanner

IOAN IACOB • CRISTIAN R. GHEORGHIU

ELEMENTE DE STATISTICĂ APLICATE ÎN EDUCAȚIE FIZICĂ ȘI SPORT

Scanned with OKEN Scanner

REFERENȚI

PROF.UNIV.DR. **ELISABETA JABA**
ȘEF AL CATEDREI DE STATISTICĂ
FACULTATEA DE ȘTIINȚE ECONOMICE
UNIVERSITATEA AL.I.CUZA IAȘI

CONF.UNIV.DR. **IOAN APOSTOL**
DIRECTORUL POLICLINICII PENTRU SPORTIVI IAȘI

ISBN 973-9312-16-0
© EDITURA UNIVERSITĂȚII "AL.I.CUZA" IAȘI
STR. PĂCURARI 9, IAȘI 6600, TEL. 032-114947

Scanned with OKEN Scanner

IOAN IACOB CRISTIAN R. GHEORGHIU

ELEMENTE DE STATISTICĂ APLICATE ÎN EDUCAȚIE FIZICĂ ȘI SPORT

EDITURA UNIVERSITĂȚII "AL.I.CUZA"
IAȘI – 1998

Scanned with OKEN Scanner

Descrierea CIP a Bibliotecii Naționale
IACOB, IOAN

Elemente de statistică aplicate în educație fizică și sport / Ioan Iacob, Cristian R. Gheorghiu – Iasi: Editura Universității "Al.I.Cuza", 1998

152 p. : il.; 21 cm.

Bibliogr.

ISBN 973-9312-16-0

I. Gheorghiu, Cristian R.

311:796

Scanned with OKEN Scanner

CAPITOLUL I

NOTIUNI GENERALE

I.1. Scurt istoric

Activitatea umană de cunoaștere a lumii înconjurătoare, de obținere a bunurilor materiale și spirituale s-a amplificat și diversificat ca urmare a perfecționării instrumentelor de cunoaștere și analiză a fenomenelor sociale și economice și a mijloacelor de producere a acestor bunuri.

Statistica ocupă un loc important în cadrul științelor sociale și are o arie foarte largă și variată de cercetare, datorită complexității și diversității fenomenelor vieții sociale.

În evoluția ei de-a lungul secolelor, statistica și-a diversificat aria de investigație, perfecționându-și, totodată, metodele sale de lucru. Drumul lung și sinuos parcurs de statistică a fost același ca și al tuturor științelor: de la necesitățile practice la elaborări teoretice.

Toate popoarele care posedă o istorie scrisă posedă și rudimente de lucrări statistice întocmite de administratorii guvernamentali în vederea rezolvării unor

Scanned with OKEN Scanner

probleme de ordin financiar, militar, fiscal, social, etc. Astfel de lucrări, în special recensăminte, destinate pentru a evidenția resursele umane și materiale ale statelor, au fost întocmite în Antichitate, în Egipt. Grecia, Roma, etc.

Prima analiză statistică, în spirit științific, a unor date, este datorată lui John Graunt (1620-1674) care, pe baza datelor deceselor înregistrate în Londra acelor ani a izbutit să tragă concluzii valabile asupra unor fenomene sociale, cum ar fi natalitatea și mortalitatea, echilibrul numeric.

Denumirea de "*statistică*" a apărut pentru prima oară în limba germană ("*statistik*") și derivă de la cuvântul latin "*status*" care se traduce prin *stat* sau *stare de lucruri*. Termenul a fost introdus în secolul al XVIII-lea de profesorul Gottfried Achenwall de la Universitatea din Göttingen.

Dezvoltarea teoriei probabilităților a constituit, fără îndoială, un mare pas înainte nu numai pentru știința statisticii, ci și pentru întreaga istorie intelectuală a omenirii. Ecuația curbei normale a fost publicată în 1733 de Abraham de Moivre, iar lucrările acestuia au fost dezvoltate ulterior în mod independent de Laplace și Gauss. Astăzi, curba normală poartă de obicei numele

Scanned with OKEN Scanner

savantului german C.F.Gauss (1775-1855), unul din savanții care au adus cele mai importante contribuții la dezvoltarea statisticii moderne, prin elaborarea metodei celor mai mici pătrate.

O contribuție importantă în direcția folosirii statisticii ca instrument de lucru în științele sociale și-a adus-o belgianul Adolphe Quételet (1796-1874) care a dat un nou impuls teoriei și practicii statisticii nu numai prin lucrările sale, ci și prin inițiativele sale organizatorice: el a organizat primul Congres Internațional de Statistică în 1853. Contribuția sa la dezvoltarea științelor sociale a constat, în special, în evidențierea variabilității fenomenelor sociale și economice de masă în jurul unor medii.

Aspectul variațional a reprezentat un concept central în teoria selecției naturale a lui Darwin ("Originea speciilor", 1859). Darwin nu a avut nici o contribuție în statistică, dar el a creat contextul teoretic, bazat pe observații, ca studiul variației să constituie o necesitate pentru rigurozitatea științifică.

Adevăratul început al statisticii moderne poate fi considerat la începutul secolului al XX-lea odată cu dezvoltarea lucrărilor lui Karl Pearson (1857-1936), care a pus bazele statisticii inductive (inferențiale) prin elaborarea testelor privitoare la semnificația diferențelor dintre valorile

Scanned with OKEN Scanner

calculate și cele empirice, toate acestea pornind de la ideeea de a conferi o bază matematică pentru teoria evoluționistă. Între anii 1894 și 1916 el a publicat 19 lucrări și monografii pe teme statistice, lucrări de o mare importanță, intitulate "Contribuții la teoria matematică a evoluției".

R. A. Fischer (1890-1962) a dezvoltat analiza variației, iar între anii 1920 și 1960 și-a dedicat timpul studierii problemelor statistice din științele biologice ("Metode statistice pentru cercetări științifice") și fundamentărilor matematice în genetică.

Deși, atât K. Pearson cât și R. A. Fischer au fost preocupați, în primul rând, de aplicarea metodelor statistice în biologie și genetică, un câmp larg de aplicare a teoriilor și metodelor lor s-a deschis, mai ales în perioada postbelică, în sociologie și economie.

În România, prima lucrare de statistică descriptivă o reprezintă "Descriptio Moldaviae" a lui Dimitrie Cantemir (1714), o monografie cu caracter geografic, politic, economic, social și cultural, care îl impune pe autorul ei printre fruntașii statistcii europene.

Scanned with OKEN Scanner

I.2. Obiectul, metoda și etapele cercetării statistice

În dezvoltarea sa, statistica s-a ocupat de acele fenomene și procese care se produc într-un număr mare de cazuri, denumite fenomene de masă sau fenomene de tip colectiv. În toate domeniile (societate, economie, natură, tehnică, etc.) fenomenele (procesele) pot apare ca rezultat al unei singure cauze sau ca rezultat a mai multor cauze care se manifestă izolat sau în interdependență.

Deși la nivelul unităților individuale se manifestă o mare variabilitate în timp și spațiu, tendința lor de manifestare nu poate fi cunoscută și verificată decât la nivelul ansamblului, la nivelul de masă al fenomenelor. Acest lucru este posibil deoarece toate fenomenele de masă se află sub incidența legii numerelor mari, potrivit căreia *variațiile întâmplătoare de la tendința centrală se compensează reciproc într-un număr mare de cazuri individuale luate în studiu.*

Obiectul de studiu al statisticii se poate defini astfel: ***statistica studiază, sub formă numerică, variația fenomenelor de masă, fenomene care sînt supuse legilor statistice, care se manifestă în condiții concrete variabile în timp și în spațiu.***

Scanned with OKEN Scanner

Statistica, și-a elaborat un ansamblu de metode științifice, care ne învață cum să organizăm observarea fenomenelor de masă și să obținem datele necesare, cum să prelucrăm aceste date și cum să formulăm ipoteze privitoare la relațiile evidențiate de aceste date. Ea oferă metode pentru testarea ipotezelor noastre și pentru confruntarea realității cu predicțiile ce le putem formula pe baza ipotezelor, permițând deci extinderea cunoașterii fenomenelor studiate.

Metodele de investigație statistică sînt reprezentate de ***totalitatea principiilor metodologice, procedeelor și tehnicilor de investigare statistică a fenomenelor ce aparțin unor procese de masă.***

Utilizarea metodelor statistice impune folosirea unui limbaj științific comun, util în raportarea la cercetări similare aceluiași fenomen sau la unele studii din discipline științifice apropiate. Metodele de investigație statistică se folosesc în studii concrete ale fenomenelor de masă și ele se diversifică în funcție de cele trei etape ale unei cercetări statistice:

- observarea statistică;
- prelucrarea statistică a datelor;
- analiza și interpretarea statistică.

Scanned with OKEN Scanner

Observarea statistică, prima etapă a oricărei cercetări statistice, constă în înregistrarea datelor, după criterii unitare, privitoare la caracteristicile colectivității supuse cercetării. Este o etapă importantă a cercetării statistice întrucât de autenticitatea datelor obținute în această etapă depinde calitatea întregii cercetări. Trebuie să se aleagă date pentru toate unitățile colectivității astfel încât, prin prelucrarea lor, să se poată desprinde regularitățile care se manifestă în cadrul colectivității;

Prelucrarea statistică reprezintă cea de a doua etapă a cercetării statistice, în care se centralizează și se sistematizează datele culese în prima etapă, se calculează sistemul de indicatori statistici cu care se caracterizează numeric fenomenele și procesele studiate și se prezintă rezultatele prelucrării sub formă de tabele, serii și grafice;

Prin prelucrarea datelor observării se asigură trecerea de la datele individuale la date și informații care permit caracterizarea întregii colectivități cercetate. Prelucrarea statistică presupune sintetizarea și compararea datelor, astfel încât să se elimine ceea ce este întâmplător și neesențial în producerea și

Scanned with OKEN Scanner

manifestarea fenomenului studiat. Prelucrările statistice reduc și elimină uneori subiectivitatea cercetătorului, contribuind la sporirea obiectivității științifice.

Procesul de investigație statistică se încheie cu analiza și interpretarea rezultatelor și cu formularea concluziilor statistice. Efectele utilizării instrumentului matematic sunt benefice pentru creșterea eficienței procesului instructiv-educativ, cu condiția ca fenomenele studiate să poată fi cuantificate, iar prelucrările statistice să fie urmate de interpretări competente pentru a deveni suportul unor intervenții ameliorative.

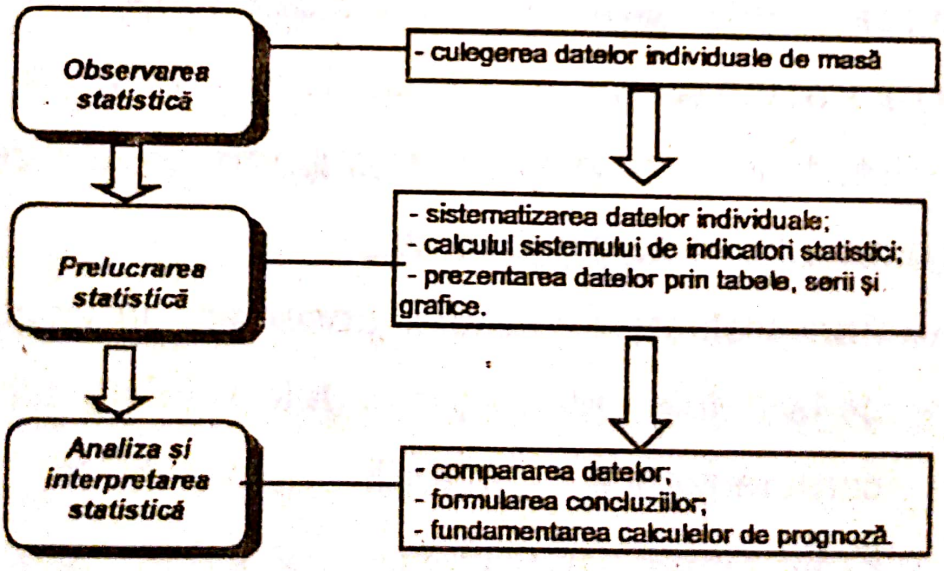


Fig. 1.1. Etapele cercetării statistice

Scanned with OKEN Scanner

I.3. Principalele noțiuni folosite în cercetarea statistică

Pentru a evita interpretările greșite în statistică și pentru a crea un cadru comun tuturor activităților investigate cu metodele statistice se utilizează termeni de bază care se regăsesc în toate cele trei etape ale cercetării statistice:

1. **Colectivitatea (populația) statistică** reprezintă totalitatea elementelor sau faptelor individuale de același fel care formează obiectul cercetării. Colectivitatea statistică are un caracter obiectiv, concret și finit. Ea trebuie delimitată precis în timp și în spațiu în cazul oricărei cercetări statistice. De exemplu: populația României, totalitatea studenților de la facultățile cu profil de educație fizică și sport, etc.

2. **Unitatea statistică** este elementul individual al colectivității. Unitățile statistice pot fi **simple** (persoana, sportivul) și **complexe** (familia, echipa, etc.). Unitățile complexe sunt formate din mai multe unități simple.

3. **Caracteristica statistică** reprezintă însușirea comună tuturor unităților colectivității. Unitățile sunt purtătoare ale unor caracteristici variabile ca nivel de dezvoltare sau formă de manifestare. Ele se mai numesc **variabile**

Scanned with OKEN Scanner

statistice. De exemplu: vârsta, sexul, etc. Se disting astfel **variabile independente** cărora li se pot atribui toate valorile cuprinse într-o anumită mulțime și **variabile dependente** ce pot fi determinate pe baza unei relații sau metode precizate în raport cu variabilele independente. De exemplu, dacă se cercetează valoarea unei metode didactice nou introdusă, aceasta reprezintă variabila independentă (cauză), iar progresul elevilor în termeni de randament școlar - variabila dependentă (variabila efect).

4. **Eșantionul** este un subansamblu al unei populații alcătuit prin extragere de elemente din populația de bază. Extragerea se face după reguli precis stabilite, în funcție de scopul analizei statistice și de natura datelor observației. Concluziile pot fi extinse apoi asupra întregii populații, bineînțeles cu o anumită marjă de eroare, care poate fi calculată.

5. **Indicatorul statistic** reprezintă expresia numerică cu un conținut real și o formă specifică de exprimare, obținută în urma efectuării unei cercetări statistice. Orice indicator statistic este format din: partea care lămurește conținutul indicatorilor, o parte care specifică timpul și spațiul la care se referă și expresia numerică. De exemplu, talia medie

Scanned with OKEN Scanner

înregistrată la școlarii clasei a IV-a (băieți) în municipiul Iași în anul 199... a avut valoarea de x cm.

6. **Datele statistice** sunt caracterizări numerice ale unităților, obținute din observare și prelucrare.

7. **Informația statistică** este mesajul datelor statistice.

8. **Frecvență sau pondere,** reprezintă numărul de unități la care se înregistrează aceeași variantă sau valoare.

9. **Seriile statistice** sunt rezultatul grupării datelor după una sau mai multe caracteristici de grupare. Dacă datele observării se grupează în funcție de o singură caracteristică rezultă serii simple. Seriile statistice de distribuție reprezintă o corespondentă între valorile caracteristicii și frecvențele unităților la care se înregistrează aceeași variantă.

Scanned with OKEN Scanner

CAPITOLUL II

DISTRIBUȚIA DE FRECVENȚE ȘI REPREZENTAREA EI GRAFICĂ

II.1. Distribuția de frecvențe

Datele obținute din desfășurarea experimentelor reprezintă de cele mai multe ori o colecție de numere. Aranjarea, clasificarea acestora pot ajuta cercetătorul să înțeleagă natura trăsăturilor esențiale a datelor recoltate și pot reduce chiar numărul calculelor matematice.

Distribuția de frecvență este reprezentată de o ordonare (aranjare) a datelor care arată frecvența de apariție a diferitelor valori ale unei variabile.

Să considerăm următorul șir de date (Tabelul 2.1) reprezentând numărul total de puncte obținute de 100 elevi după trecerea unor probe de control.

Ca prim pas este necesară scrierea (așezarea) datelor în ordine, de exemplu de la cea mai mare valoare (134) la cea mai mică (67), putându-se face ordonarea și în sens crescător . La o primă inspecție se observă că

Scanned with OKEN Scanner

anumite valori apar de mai multe ori (2 de 121, 3 de 111, 4 de 109, etc.) ceea ce ne-ar permite reașezarea datelor

Ordonarea datelor înregistrate la trecerea probelor de control

Tabel 2.1

127	109	101	92	82
121	108	100	91	82
119	107	100	91	81
117	106	99	90	80
117	105	98	90	80
114	104	96	89	79
111	104	94	87	77
111	103	93	84	75
110	103	93	84	72
109	102	93	83	67

în coloane, una conținând numărul punctelor acumulate de elevi, iar cealaltă de câte ori acest număr apare (se repetă). O astfel de rearanjare (Tabelul 2.2) poartă numele de serie de distribuție a frecvențelor, iar numărul de

Scanned with OKEN Scanner

repetări al fiecărei valori îl reprezintă frecvența (notată cu *f*).

Tabel de distribuție a frecvențelor valorilor obținute la trecerea probelor de control

Tabel 2.2

134	117	100	83
133	116	99	82
132	115	98	81
131	114	97	80
130	113	96	79
129	112	95	78
128	111	94	77
127	110	93	76
126	109	92	75
125	108	91	74
124	107	90	73
123	106	89	72
122	105	88	71
121	104	87	70
120	103	86	69
119	102	85	68
118	101	84	67

În tabelul 2.2 datele sunt ordonate în tot atâtea clase câte valori sînt (de la 134 la 67), respectiv 68 de clase, ceea ce reprezintă un număr destul de mare a acestora. De obicei se recurge la reducerea numărului de clase prin aranjarea datelor în grupuri; de pildă, cifrele de

Scanned with OKEN Scanner

la 134 la 130 sau de la 80 la 84 pot fi grupate împreună, așa cum este redat în tabelul 2.3.

Tabel de distribuție al frecvențelor valorilor obținute la trecerea probelor de control

Tabel 2.3

Clase	
130-134	
125-129	
120-124	
115-119	
110-114	
105-109	
100-104	
95-99	
90-94	
85-89	
80-84	
75-79	
70-74	
65-69	
Total	

Diferența dintre tabelele 2.2 și 2.3 este dată de intervalul de clasă ales, care este 1 pentru primul tabel și 5 pentru cel de-al doilea precum și de numărul de clase care este 68, respectiv 14. În felul în care sînt aranjate datele în tabelul 2.2 cu intervalul de clasă egal cu 1, șirul original de date (ordonate) poate fi reconstituit ușor după

Scanned with OKEN Scanner

frecvență, fără a pierde vre-o informație. Dar dacă intarvalul de clasă ales este mai mare decât 1 o parte din valorile individuale se vor pierde, deci șirul original nu va putea fi reconstituit exact după frecvențe, iar pierderea unor informații (date) este inevitabilă. Ca urmare s-au impus câteva convenții pentru alegerea numărului de clase și a mărimii intervalului de clasă:

1. numărul de clase să fie de minim 10 și de maxim 20 de clase. De exemplu, dacă valoarea minimă dintr-un șir de observații este 7, iar cea maximă 156, un interval de clasă ales de 10 va conduce la obținerea unui număr de 16 clase. Dacă valoarea minimă este 2, iar cea maximă este 38, putem obține 14 clase (cu intervalul de clasă de 3). Dacă într-un șir minimul este 9 iar valoarea maximă este 20, un interval de clasă de 1 va conduce, așa cum ar fi recomandabil, la 12 clase.
2. intervalul de clasă ales să fie de 1, 2, 3, 5, 10 sau 20. O formulă des utilizată pentru calcularea intervalului de clasă este:

$$i = \frac{1}{100} \times 8(x_{max} - x_{min}) \text{ ,unde}$$

i = intervalul de clasă;

X_{max} = valoarea maximă din șir;

X_{min} = valoarea minimă din șir.

Scanned with OKEN Scanner

Pentru datele din tabelul 2.1, de exemplu, avem:

$$i = \frac{1}{100} \times 8(134 - 67) = \frac{8 \times 67}{100} = 5{,}36$$

deci rotunjit 5, așa cum am procedat și exemplificat în tabelul 2.3.

3. pe cât posibil, intervalul de clasă să fie un număr impar, iar centrul intervalului de clasă (despre care vom vorbi imediat) să fie un număr întreg.
4. valorile de început ale claselor este recomandabil să fie multiplii ai intervalului de clasă. De exemplu, pentru un interval de clasă de 5, clasele încep cu 5, 10, 15, 20, etc. (în tabelul 2.3 - 130, 125, 120, 115 ... sînt multiplii de 5).
5. aranjați clasele în ordine descrescătoare, plasând valorile cele mai crescute la începutul (partea de sus) coloanei.

Pentru calcularea diferiților indicatori statistici sau pentru a reprezenta grafic o distribuție de frecvențe este necesară precizarea a încă două aspecte importante.

Primul se referă la faptul că datele din interiorul unei clase sînt uniform distribuite între limitele exacte

Scanned with OKEN Scanner

(reale) ale intervalului ales. De exemplu, pentru înregistrarea vitezei de reacție la un semnal s-a înregistrat și valoarea de 0,196 secunde. Convențional, și pentru a conferi acuratețe măsurării, se consideră că limitele exacte ale valorii se află cu o jumătate de unitate în plus și în minus față de valoarea observată, în cazul nostru 0,1955 s și 0,1965. Pentru datele din tabelul 2.4 se observă că pentru clasa 100-104 limitele exacte ale intervalului sunt 99,5 și 104,5 cu frecvența de 16.

Clasele, limitele exacte și valorile centrale ale clasei pentru probele de control (viteza de reacție)

Tabel 2.4

Clase	Valoarea centrală a clasei
130-134	132
125-129	127
120-124	122
115-119	117
110-114	112
105-109	107
100-104	102
95-99	97
90-94	92
85-89	87
80-84	82
75-79	77
70-74	72
65-69	67

Scanned with OKEN Scanner

Așa cum afirmam, cele 16 valori sunt distribuite uniform (egal) după cum urmează:

103,5 - 104,5	
102,5 - 103,5	
101,5 - 102,5	
100,5 - 101,5	
99,5 - 100,5	
Total	

Aceasta ne va folosi la calcularea medianei și la reprezentarea histogramelor.

Al doilea aspect se referă la faptul că se consideră că toate datele din interiorul unei clase se concentrează în jurul valorii centrale a acesteia. Valoarea centrală a unei clase se află adunând cele două valori extreme ale clasei și împărțind rezultatul la doi. De exemplu, pentru clasa aleasă de noi 100 - 104, valoarea centrală este:

$$\frac{100+104}{2}=\frac{204}{2}=102$$

sau ținând seama de limitele exacte ale clasei

$$\frac{(99,5+104,5)}{2}=102$$

Scanned with OKEN Scanner

Aceasta ne va folosi la calcularea mediei aritmetice, a abaterii standard și la reprezentarea poligonului de frecvențe.

II.2. Reprezentarea grafică a distribuției

În unele cazuri, o prezentare grafică a tabelului de frecvențe ne oferă informații mai clare și mai concise despre o distribuție de frecvență. Există mai multe feluri de astfel de prezentări, rezumându-ne în lucrarea de față la cele cu cea mai largă utilizare: histograma, poligonul frecvențelor și curba cumulativă.

Folosind tabelul 2.5 de frecvență a înregistrărilor timpului de reacție auditivă la un număr de 188 de subiecți să trasăm histograma corespunzătoare.

Pe abscisa (Ox) coordonatelor ortogonale, vom marca limitele exacte (reale) ale claselor. Pe ordonată (Oy) vom trece punctele corespunzând frecvențelor din fiecare grupă. Ridicăm apoi pe abscisă dreptunghiuri cu baza egală și cu înălțimea corespunzătoare frecvențelor din fiecare grupă. Primul dreptunghi va avea înălțimea de 1, corespunzătoare numărului de frecvențe din prima

Scanned with OKEN Scanner

Distribuţia de frecvenţă a timpului de reacţie

Tabel 2.5

Limitele reale (exacte)	Valoarea centrală a clasei	Frecvenţe cumulate
0,335 - 0,355	0,345	188
0,315 - 0,335	0,325	186
0,295 - 0,315	0,305	184
0,275 - 0,295	0,285	180
0,255 - 0,275	0,265	175
0,235 - 0,255	0,245	164
0,215 - 0,235	0,225	147
0,195 - 0,215	0,205	119
0,175 - 0,195	0,185	50
0,155 - 0,175	0,165	13
0,135 - 0,155	0,145	1

Scanned with OKEN Scanner

clasă, al doilea de 12 și așa mai departe încât obținem:

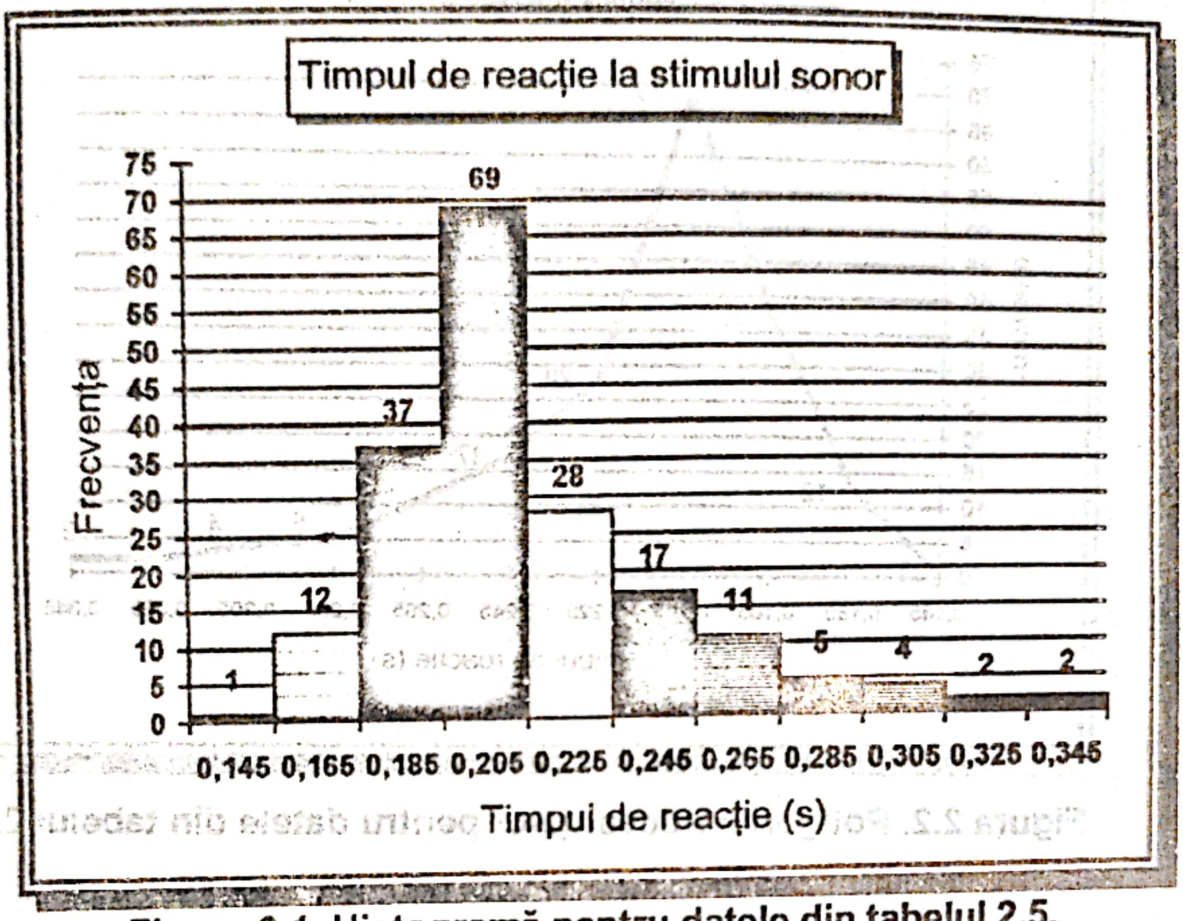


Figura 2.1. Histogramă pentru datele din tabelul 2.5.

Dacă unim punctele reprezentând valoarea centrală a claselor cu linii drepte obținem un poligon al frecvențelor, așa cum se observă în figura 2.2.

Se remarcă faptul că distribuția de frecvență reprezentată în figura 2.2. nu este o linie continuă, ci una frântă în fiecare punct de intersecție al valorilor centrale ale clasei cu frecvența corespunzătoare.

Scanned with OKEN Scanner

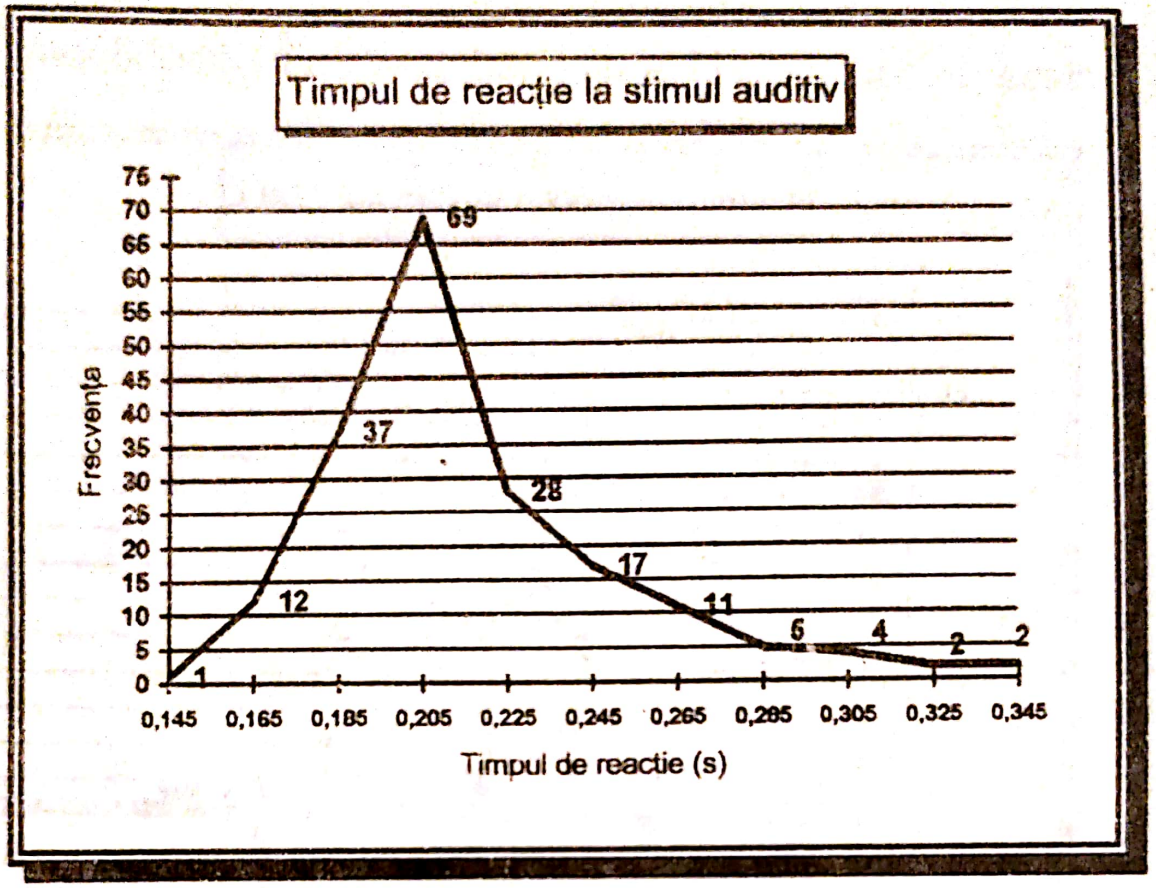


Figura 2.2. Poligonul frecvențelor pentru datele din tabelul 2.5.

Trasarea curbei cumulative se aseamănă cu cea a poligonului de frecvență, diferind de acesta prin două aspecte:

1. se folosesc frecvențele cumulate în locul celor simple;

2. de pe abscisă se ridică verticala corespunzătoare limitei superioare exacte a clasei și nu a valorii centrale (mijlocul) a acesteia. Se procedează astfel deoarece se urmărește ca graficul nostru să redea exact numărul de cazuri de deasupra sau dedesubtul unei valori.

Scanned with OKEN Scanner

Curbele frecvențelor cumulate permit localizarea în cadrul intervalului total de variație a claselor cu frecvențe mari (panta mai accentuată) sau mici (pante mai reduse). Importanța lor rezultă însă din faptul că ele ne oferă răspunsul rapid la două întrebări:

- ✓ care este proporția (probabilitatea) cazurilor cărora le corespunde o valoare inferioară sau superioară unei anumite valori date (prag);
- ✓ invers, care este valoarea căreia îl corespunde o proporție (probabilitate) dată a cazurilor cu valori inferioare sau superioare modalității (valorii) respective?

Să trasăm mai întâi curba cumulativă ținând cont de frecvențele cumulate din tabelul 2.5.

Acum putem da drept exemplu. un răspuns la prima întrebare. Care este proporția (probabilitatea) cazurilor care au reacționat la stimulul sonor în mai mult de 0,255s? Ridicând verticala de pe abscisă aceasta întâlnește curba în punctul corespunzător ordonatei la valoarea frecvenței cumulate 164. Dacă 188 reprezintă 100% din totalul cumulat, 164 va reprezenta 164/188x100, respectiv 87,234%. Deci 87,234 dintre subiecți au răspuns la stimul într-un timp mai scurt de 0,255s, iar diferența de

Scanned with OKEN Scanner

100 - 87,234 respectiv 12,7659%, într-un timp mai îndelungat, aceasta fiind soluția problemei noastre.

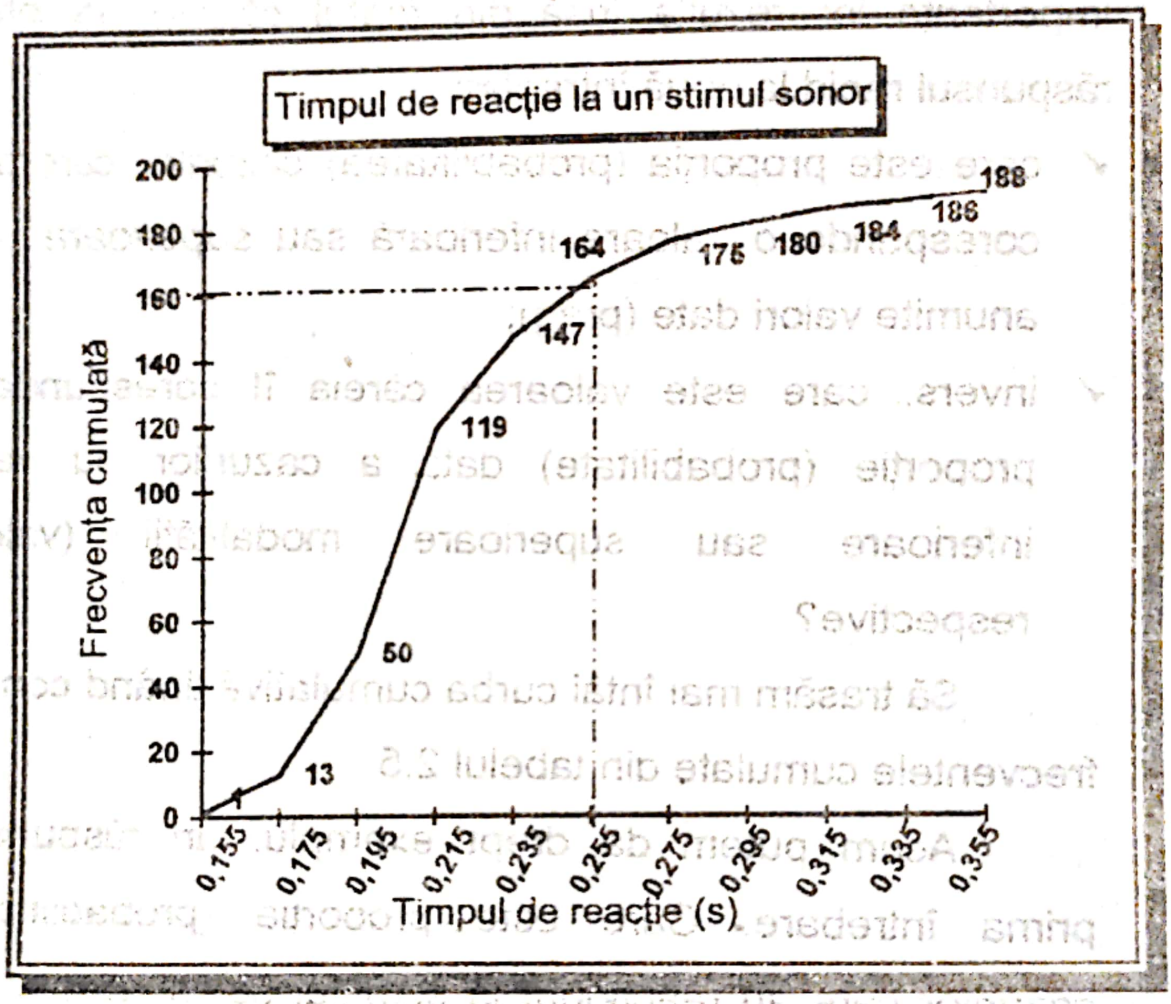


Figura 2.3. Curba cumulativă pentru datele din tabelul 2.3.

Invers, cunoscând proporția, putem afla timpul de reacție și obține răspunsul la cea de-a doua întrebare.

Dacă am considera distribuția aceasta cu o infinitate de clase și trecând de la noțiunea de proporție la cea de probabilitate, putem spune că probabilitatea de a

găsi o valoare inferioară celei de 0,255 este de 0,872 (sau 87,234%) și de a găsi una superioară este 0,127.

Exerciții

1. Cunoscând valoarea minimă și maximă dintr-un set de observații, să se stabilească intervalul de clasă, limitele intervalului cel mai mic, limitele exacte ale clasei și valoarea centrală a acesteia.

4 - 39
35 - 96
0,141 - 0,205

2. Se dă următorul set de observații:

48	64	57
65	60	63
57	47	57
53	61	70
67	60	81
47	62	81
53	51	60
66	76	52
71	73	50
47	54	69

Scanned with OKEN Scanner

Se cere să se realizeze o distribuție de frecvențe, considerând intervalul de clasă de 5.

Răspunsuri:

1.

10	80 - 89	79,5 - 89,5	84,5
1	17	16,6 - 17,5	17
20	0 - 19	-0,5 - 19,5	9,5

2. Clasa cu valorile cele mai mici din șir va fi 35-39, 40-44, etc. și vom avea frecvențele:

5, 4, 4, 8, 11, 12, 11, 6, 2, 1.

Scanned with OKEN Scanner

CAPITOLUL III

PARAMETRII TENDINȚEI CENTRALE

Datele obținute în urma observării caracterizează fiecare unitate a colectivității care urmează să fie studiată și sunt expresia acțiunii unui număr mare de cauze și factori care pot acționa în același sens sau în sensuri diferite. Cercetarea statistică presupune însă cunoașterea trăsăturilor reprezentative esențiale ale fenomenelor de masă, a regularităților de producere a acestora. Aceasta înseamnă că este necesară determinarea unor valori reprezentative, tipice pentru masa valorilor individuale. Pentru caracterizarea numerică sintetică a unei distribuții statistice, implicit a curbei de distribuție corespunzătoare, se determină (în general) trei tipuri de parametrii (indici), fiecare tip descriind anumite caracteristici ale acestora:

1. parametrii tendinței centrale, ne dau informații despre zona de maximă concentrare a valorilor (zona frecvențelor maxime), astfel încât în jurul lor se plasează, la distanțe mai mari sau mai mici, celelalte valori din șir;

Scanned with OKEN Scanner

2. parametrii dispersiei, care exprimă gradul de împrăștiere a valorilor din șir în jurul valorilor centrale;
3. parametrii formei (curbei de distribuție), care oferă o măsură a asimetriei împrăștierii valorilor distribuției față de tendința centrală.

Parametrii (indicatorii) tendinței centrale cei mai frecvent utilizați în analizele statistice sînt

- ✓ **media**,
- ✓ **mediana** și
- ✓ **modulul**.

III.1. Media aritmetică simplă și ponderată

Există, de fapt, mai multe feluri de medii. Cea mai des întâlnită este media aritmetică, denumită pe scurt, media. Ea este rezultatul sintetizării printr-o singură expresie numerică a tuturor valorilor individuale înregistrate, obținută prin raportarea valorii totalizate ale caracteristicii la numărul total al unităților. Calculată pentru o colectivitate statistică, media aritmetică este valoarea la care ar fi ajuns fiecare unitate înregistrată dacă toți factorii ar fi acționat în mod constant în toate cazurile.

Scanned with OKEN Scanner

III.1.1. Media aritmetică simplă

Media aritmetică se obține deci prin însumarea elementelor unei serii statistice și prin împărțirea totalului la numărul elementelor care formează seria.

Ea are în statistică două simboluri:

a) $\overline{X}$ - pentru media aritmetică a unei populații;

b) $\overline{x}$ - pentru media aritmetică a unui eșantion.

Deosebirea dintre cele două simboluri este importantă, deoarece media aritmetică a unei populații este o mărime fixă (deși nu o cunoaștem) înainte de a o calcula, pe când media aritmetică a eșantioanelor poate varia la fiecare dintre acestea, dintr-o populație putând fi extrase mai multe eșantioane care vor avea medii aritmetice diferite.

Să considerăm următorul șir de date: 7, 13, 22, 9, 11, 4. Conform definiției, suma acestora este 7+13+22+9+11+4 = 66, iar numărul elementelor seriei este 6; în concluzie, media aritmetică simplă este 11 (66/6).

În termeni algebrici, dacă notăm prima observație cu x_1, a doua cu x_2, și cu x_n ultima observație, iar cu N numărul observațiilor, vom avea pentru întreaga populație:

Scanned with OKEN Scanner

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{N} = \frac{\sum_{i=1}^{N} Xi}{N}$$

iar pentru un eșantion:

$$\bar{x} = \frac{x_1 + x_2 + x_3 \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

În cazul nostru :

$$\frac{7+13+22+9+11+4}{6} = \frac{66}{6} = 11$$

Calculul mediei aritmetice este, după cum se vede, foarte simplu.

III.1.2.Media aritmetică ponderată

Când ne confruntăm cu o distribuție de frecvențe, pentru calcularea mediei aritmetice, nu mai putem aduna observațiile individuale, fiindcă observațiile au fost grupate în clase.

Așa încât vom recurge la calculul mediei aritmetice ponderate raportând suma produselor dintre valorile caracteristicii (x_i) și frecvența lor de apariție (f_i) la totalul frecvențelor :

Scanned with OKEN Scanner

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \ldots + x_n f_n}{\sum f_i} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum f_i}$$

De exemplu, pentru datele înregistrate la aruncarea mingii medicinale (2kg) înapoi cu ambele mâini, la un colectiv de elevi s-au înregistrat următoarele performanțe (Tabelul 3.1):

Calcularea mediei ponderate la aruncarea mingii medicinale înapoi

Tabelul 3.2

18	1	18
16	2	32
14	2	28
12	3	36
Total	**N = 20**	**Σ = 280**

Rezultă că $\bar{x} = 280/20 = 14$ m.

Pentru distribuții de frecvență în care intervalul de clasă este mai mare decât 1, vom folosi drept valoare a caracteristicii valoarea centrală a clasei. Să exemplificăm cu datele din Tabelul 3.3:

Scanned with OKEN Scanner

Calculareа mediei ponderate
(pentru intervalul de clasă mai mare ca 1)

Tabel 3.3

	47		47
	42		84
	37		111
	32		192
	27		216
	22		374
	17		442
	12		132
	7		14
	2		0

$$\bar{x} = \frac{1612}{76} = 21{,}21$$

Mai explicit:

pas 1: am calculat valoarea centrală a clasei (vezi II.1);

pas 2: am înmulțit valoarea centrală (x_i) cu frecvența (f_i) corespunzătoare (coloana f_ix_i);

pas3 : am adunat produsele obținute (1612);

pas 4: adunăm valorile din coloana frecvențelor f_i (76);

Scanned with OKEN Scanner

pas 5: aplicând formula de calcul a mediei ponderate am împărțit suma produselor $f_i x_i$ la totalul frecvențelor (76).

III.1.3 Proprietațile mediei aritmetice

Media aritmetică are o serie de proprietăți, dintre care le menționăm pe cele mai importante:

1. mărimea mediei aritmetice este întotdeauna cuprinsă în intervalul de variație al valorilor individuale ale caracteristicii:

$$x_{min} < \bar{x} < x_{max},$$

unde x_{min} = valoarea cea mai mică din șir;

x_{max} = valoarea cea mai mare din șir.

2. Suma algebrică a abaterilor termenilor seriei de la media lor este egală cu zero.

În cazul unei serii simple aceasta înseamnă :

$$\Sigma(x_i - \bar{x}) = 0,$$

iar în cazul unei serii cu frecvență

$$\Sigma(x_i - \bar{x})f_i = 0.$$

Scanned with OKEN Scanner

Mai explicit, considerând următorul șir 7, 13, 22, 9, 11, 4 avem media egală cu 66/6 = 11. Deci:

		$\bar{x}$	$x_i - \bar{x}$
	-	11	= -4
	-	11	= 2
	-	11	= 11
	-	11	= -2
	-	11	= 0
	-	11	= -7

3. Suma pătratelor abaterilor dintre date și media aritmetică este mai mică decât suma abaterilor ridicate la pătrat în jurul oricărei alte valori în afară de media aritmetică. Folosind exemplul anterior se poate ușor demonstra această proprietate:

	$\bar{x}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	11	-4	16
	11	2	4
	11	11	121
	11	2	4
	11	0	0
	11	-7	49
			$\Sigma(x_i - \bar{x}) = 194$

Dacă alegem o altă valoare decât media aritmetică (să zicem 4), suma abaterilor ridicate la pătrat va fi mai mare

decât 194: $(7-4)^2 + (13-4)^2 + (22-4)^2 + (9-4)^2 + (11-4)^2 + (4-4)^2 = 488$. este evidentă relația: $488 > 194$.

4. La seriile de distribuție de frecvență, media oscilează în jurul termenului cu frecvența cea mai mare.

5. Într-un șir de valori egale, media acestora este egală cu fiecare dintre ele, indiferent dacă este o serie simplă sau cu frecvențe. Dacă avem șirul 3, 3, 3, 3, 3, 3, media aritmetică este 3, respectiv ca fiecare valoare a șirului dat.

Câteva caracteristici ale mediei aritmetice:

- ✓ calculul mediei este simplu;
- ✓ există o singură medie aritmetică pentru un ansamblu dat de valori;
- ✓ media ia în considerare toate valorile ansamblului;
- ✓ trebuie să fie exprimată în aceleași unități de măsură ca și valorile la care se referă (metru, secundă, greutate, puncte, etc.).

Dezavantaje:

- deoarece ia în considerare toate valorile ansamblului devine sensibilă la valorile foarte mari sau foarte mici;

Scanned with OKEN Scanner

astfel media nu mai reflectă corect tendința centrală a distribuției;

- media "comprimă" excesiv valorile extreme, comparativ cu cele aflate mai spre centrul distribuției; la distribuțiile simetrice această situație contează mai puțin, însă la cele puternic asimetrice acest lucru nu este de neglijat.

III.1.4. Media aritmetică pentru mai multe șiruri

Pentru două sau mai multe șiruri (k), media aritmetică se calculează prin împărțirea sumei produselor dintre fiecare medie și numărul de cazuri din grup la numărul total de cazuri:

$$\bar{x} = \frac{\sum_{i=1}^{n_1} x_{i1} + \sum_{i=1}^{n_2} x_{i2}}{n_1 + n_2} = \frac{n_1 x_1 + n_2 x_2}{N}$$

$$\bar{x} = \frac{n_1 x_1 + n_2 x_2}{n_1 + n_2} = \frac{4 \times 5 + 6 \times 7}{4 + 6} = \frac{20 + 42}{10} = 4{,}2$$

Scanned with OKEN Scanner

III.2 Media armonică simplă și ponderată

III.2.1. Media armonică simplă

Media armonică este inversul mediei aritmetice a inversurilor valorilor a caror medie o căutăm. Este un tip de medie mai rar utilizată, dar care are aplicabilitate în domeniul educației fizice și sportului, mai ales pentru rezolvarea probiemelor legate de viteze.

Să presupunem că într-o competiție aviatică se cere efectuarea unui raid aerian, fiecare avion având de parcurs un traseu de formă pătratică, cu laturile de 100 km. Unul dintre concurenți parcurge cele 4 laturi cu următoarele viteze: 100 km/h (a zburat cu vântul din față), a doua cu 200 km/h, a treia cu 300 km/h și ultima cu 400 km/h. Ne întrebăm care a fost viteza medie a avionului?

Recurgând la calculul mediei aritmetice avem:

$$\frac{100+200+300+400}{4}=\frac{1000}{4}=250km/h$$

Este rezultatul corect? Să ne convingem raționând în felul următor:

- prima latură a fost parcursă într-o oră;
- a doua în 30 de minute;
- a treia în 20 de minute;

Scanned with OKEN Scanner

- ultima în 15 minute.

În total 60 + 30 + 20 + 15 minute = 2 ore și 5 minute. Deci, viteza medie a fost în realitate alta.

Formula mediei armonice este inversul mediei aritmetice a inversului valorilor:

$$\bar{x}_h = \frac{n}{\sum \frac{1}{x}}$$

În cazul avionului participant la raid avem, deci:

$$\bar{x}_h = \frac{4}{\frac{1}{100}+\frac{1}{200}+\frac{1}{300}+\frac{1}{400}} = \frac{4}{\frac{25}{1200}} = \frac{4\times 1200}{25} = 192 km/h$$

Media armonică ne dă răspunsul corect la problema noastră. Se observă că media armonică se aplică în acest caz deoarece duratele au fost diferite pe parcursul de lungimi egale. Dacă duratele ar fi fost identice și parcursurile diferite, s-ar fi aplicat media aritmetică.

III.2.2. Media armonică ponderată

Media armonică ponderată se obține pe baza produselor dintre frecvențe și valorile inverse ale caracteristicii, astfel:

Scanned with OKEN Scanner

$$\bar{x}_h = \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} \frac{1}{x_i} f_i}$$

Media armonică calculată pe baza acestei relații, din valori pozitive, este întotdeauna mai mică decât media aritmetică.

III.3. Mediana și modulul

III.3.1. Mediana(M_e)

Orice medie este reprezentativă pentru grupuri omogene de date. Există însă multe distribuții care deși sunt alcătuite din elemente omogene prezintă mari diferențe între valorile lor extreme, diferențe agravate de faptul că elementele lor au tendința de a se concentra spre una din extreme. Astfel de distribuții sunt, într-un grad mai mare sau mai mic, asimetrice. Rezultă doar că, în astfel de cazuri, avem nevoie de o măsură a tendinței centrale care să nu fie influențată de valorile extreme. Mediana este o astfel de măsură a tendinței centrale.

Scanned with OKEN Scanner

Mediana este valoarea centrală a unei serii, ordonată crescător sau descrescător și care împarte termenii seriei în două părți egale: 50% din termeni (jumătate) sunt mai mici decât mediana și 50% sunt mai mari decât aceasta.

Dacă seria de date este impară, mediana va fi acea valoare care corespunde celei de-a (N + 1)/2 observații. Pentru seria 2, 7, 16, 19, 20, 25 și 27 (N = 7), mediana este a patra valoare, respectiv 19.

Dacă seria are un număr par de date, mediana se calculează ca medie aritmetică a celor două valori din mijlocul șirului. Să adugăm la seria anterioară valoarea 31 (N = 8); mediana va fi media aritmetică dintre 19 și 20 (19 + 20)/2, deci 19,5.

O situație mai aparte o constituie seriile mici (n<30), în care anumite valori apar de mai multe ori și nu este necesară gruparea datelor în clase. De exemplu, șirul 7, 7, 7, 8, 8, 8, 9, 9, 10, 10 (N=10) cu număr par de date. Care este mediana? În acest caz ea se obține prin interpolare, limitele exacte (vezi cap.II.1) sînt cuprinse între 7,5 și 8,5 și fiecare dintre cei trei de 8 ocupă o treime din interval (0,33). Deci mediana va fi 7,5 + 0,66 = 8,16.

Scanned with OKEN Scanner

În cazul unei distribuții de frecvențe găsirea medianei este ceva mai dificilă; astfel putem presupune că o vom afla în interiorul acelei clase care conține valoarea centrală.

Să considerăm datele din Tabelul 3.4, conținând numărul de puncte obținute de o grupă de elevi la un complex tehnico-tactic:

Distribuția de frecvențe a punctelor obținute la complexul tehnico-tactic

Tabel 3.4

45 - 49		76
40 - 44		75
35 - 39		73
30 - 34		70
25 - 29		64
20 - 24		56
15 - 19		39
10 - 14		13
5 - 9		2
0 - 4		0

Formula de calcul a medianei este:

Scanned with OKEN Scanner

$$M_e = x_0 + \left(\frac{\frac{\sum f_i + 1}{2} - \sum f_c}{f_m} \right) K,$$

cu următoarele semnificații:

x_0 - limita inferioară a clasei care conține mediana. Având 76 de observații (număr par) mediana va fi în clasa care conține a 38-a valoare. În cazul nostru a 38-a valoare este, conform frecvențelor cumulate în clasa 15-19, a cărei limită inferioară este 14,5;

Σf_i - numărul de cazuri (total 76);

Σf_c - suma cumulată a frecvențelor din clasele ce se află înaintea celei în care se găsește mediana (13 în cazul nostru);

f_m - frecvența grupei în care se găsește mediana (26);

K - intervalul de clasă (5).

În final avem:

$$M_e = 14,5 + \left(\frac{\frac{76+1}{2} - 13}{26} \right) \times 15 = 19,4$$

III.3.2 Modulul (M_o)

Modulul (modul sau dominantă) reprezintă valoarea caracteristicii cu frecvența cea mai mare. Desigur că un

colectiv este mai bine caracterizat de o valoare înregistrată care apare de mai multe ori, decât de o alta care apare o singură dată.

Să considerăm următoarele observații: 11, 11, 12, 12, 12, 13, 13, 13, 13, 13,14, 14, 14, 15, 15, 15, 16, 16, 17, 18. Valoarea 13 apare de 5 ori, mai frecvent decât oricare altă valoare; deci, modulul este 13.

În situația în care toate valorile apar cu aceeași frecvență sau nici una din valori nu se repetă, nu există modul. În următorul set de observații: 2, 7, 16, 19, 20, 25, 27 nu există modul. Nici în următoarea situație modulul nu poate fi determinat: 2, 2, 2, 7, 7, 7, 16, 16, 16, 19, 19, 19, 20, 20, 20 - toate valorile apar de trei ori.

În cazul în care două valori alăturate au aceeași frecvență, dar care este superioară frecvențelor de apariție a tuturor celorlalte valori, modulul se obține din media aritmetică a valorilor alăturate. De exemplu în seria 11, 11, 12, 12, 12, 13, 13, 13, 13, 14, 14, 14, 14, 15, 15, 16, 16, 17, 18, valorile 13 și 14 apar de 4 ori, fiind totodată frecvența cea mai mare; modulul este deci (13 + 14)/2 = 13,5.

Când două valori <u>nu</u> sunt alăturate dar apar cu frecvențe mai mari decât cele din vecinătatea lor, atunci

Scanned with OKEN Scanner

oricare din aceste două valori poate fi modul: 11, 11, 12, 12, 12, 13, 13, 13, 13, 13, 14, 14, 14, 15, 15, 15, 15, 16, 16, 16, 17, 17, 18. Valoarea 13 apare de 5 ori, mai frecvent decât 12 și 14; valoarea 15 apare de 4 ori, de mai multe ori decât 14 și 16. Mai concret, măsurarea taliei unui eșantion de bărbați și de femei va duce la concentrarea valorilor obținute la bărbați într-un alt punct decât cel al taliei femeilor. Avem de-a face în acest caz cu o distribuție de frecvență bimodală. În cazul distribuțiilor bi- sau plurimodale vom avea un mod principal și unul, respectiv mai multe moduri secundare.

În cazul distribuțiilor cu clase de frecvențe se vorbește despre o clasă modală, respectiv clasa cu frecvența cea mai mare. Modulul se calculează cu ajutorul unei formule care ia în considerare atât clasa modală, cât și clasele vecine acesteia, formulă bazată pe principiul interpolării (vezi și mediana):

$$M_0 = x_0 + \left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right) \times K \text{, unde}$$

x_0 - limita inferioară a clasei care conține modulul, clasa cu cea mai mare frecvență (dacă există două asemenea grupe, formula nu mai este valabilă);

Δ_1 - diferența dintre frecvența clasei modale și a clasei imediat următoare;

Scanned with OKEN Scanner

Δ_2 - diferența dintre frecvența clasei modale și frecvența grupei precedente;

K - intervalul de clasă.

Prin felul în care se determină, modulul nu posedă calități aritmetice bine definite și deci nu poate fi preluat în calculele ulterioare pentru determinarea unor caracteristici suplimentare ale distribuțiilor statistice (de frecvențe).

Exerciții

1. Să se calculeze media aritmetică, mediana și modulul pentru următoarele date:

1.	
2.	
3.	
4.	
5.	

2. Să se calculeze media aritmetică, mediana și modulul pentru următoarele date:

Clase	
52-53	
50-51	
48-49	
46-47	
44-45	
42-43	
40-41	
38-39	
36-37	
34-35	

Scanned with OKEN Scanner

32-33

Răspunsuri:

1.

	1.		3.		5.
	11		9,6		15,6
	11		9,1		16
			9		

2.

	41,7
	42,2
	42,5

Scanned with OKEN Scanner

CAPITOLUL IV

PARAMETRII DISPERSIEI

Prin calcularea mediei se urmărește găsirea unei expresii numerice sintetice, reprezentativă pentru toate valorile individuale din care s-a calculat. Ea va fi cu atât mai reprezentativă cu cât valorile individuale sunt mai apropiate între ele. Ca urmare, folosirea mediei în vederea caracterizării a ceea ce este comun și tipic în formele individuale trebuie să fie precedată de verificarea gradului de reprezentativitate al acesteia pentru întrega masă de valori individuale. Aceasta presupune să se analizeze în ce măsură valorile individuale variază în jurul mediei.

Dacă media este reprezentativă, semnificativă pentru valorile din care s-a calculat, atunci colectivitatea este omogenă și deci abaterea valorilor individuale de la medie poate fi explicată pe seama acțiunii factorilor întâmplători. În schimb, dacă variația în jurul mediei depășește un anumit prag, înseamnă că cel puțin unul din factorii esențiali are influență variabilă, structurând colectivitatea pe mai multe tipuri calitative.

Scanned with OKEN Scanner

Reprezentând datele grafic, cea mai simplă și mai eficientă indicație asupra împrăștierii valorilor în cadrul intervalului total de variație o oferă curba de distribuție a frecvențelor. Pentru exprimarea numerică a acestei împrăștieri se folosesc parametrii dispersiei.

La determinarea acestor parametrii se pot lua în considerare valorile extreme, mediana sau media aritmetică.

IV.1. Amplitudinea de variație(A)

Amplitudinea (A) se calculează ca diferență între valoarea maximă și valoarea minimă a caracteristicii și se exprimă în unitatea de măsură a acesteia:

$$A = x_{max} - x_{min}$$

Următoarea serie 10, 12, 15, 18, 20 are amplitudinea A=10 (obținută prin 20 -10).

În cazul datelor grupate amplitudinea variației se calculează ca diferența între limita superioară a ultimului și limita inferioară a primului interval.

Scanned with OKEN Scanner

Fiind calculată numai pe baza valorilor extreme amplitudinea nu este un indicator suficient de semnificativ pentru măsurarea variației. Se folosește mai frecvent, în prelucrarea datelor, la alegerea numărului de grupe (clase) și a mărimii intervalului de grupare.

IV.2. Cuartilele și abaterile cuartilice

Dacă valoarea centrală este mediana, atunci dispersia poate fi exprimată sub forma abaterilor cuartile.

Cuartilele (sau quartile) sunt parametrii analogi medianei, dar care împart șirul de valori în patru părți egale, fiecare incluzând 25% din valorile șirului. Atunci când se dorește o analiză mai detaliată a modului cum sunt repartizate valorile în interiorul intervalului total de variație, șirul poate fi împărțit în mai mult de 4 părți egale. Valorile care delimitează aceste subșiruri sunt denumite generic cuantile (exemple de cuantile: mediana, cuartile, decile - împart șirul în 10 părți egale -, centile, etc.).

De-a lungul unui șir ordonat crescător vom întâlni deci un număr de 3 cuartile, notate de regulă Q_1, Q_2 (coincide cu mediana) și Q_3. Astfel se pot calcula doi indici de dispersie:

Scanned with OKEN Scanner

a) abaterea cuartilică (intercuartilică), care reprezintă diferența, în valoare absolută, dintre a trei și prima cuartilă (Q_3-Q_1). Se obține astfel o imagine asupra dispersiei, comparând abaterea intercuartilică cu amplitudinea (A);

b) abterea semiintercuartilică (Q) este definită ca fiind media probabilă a abaterilor de la mediană, cu alte cuvinte jumătate din modalitățile caracteristicii diferă de mediană prin mai mult decât această valoare, iar cealaltă jumătate prin mai puțin:

$$Q = \frac{(Q_3 - Q_2)}{2}$$

Deși prin măsurarea abaterii semiintercuartilice evităm influența valorilor extreme, pierdem în același timp informațiile oferite de valorile din primul și ultimul sfert al seriei; se neglijează astfel valorile situate sub prima cuartilă sau peste cea de-a treia, deci 50% din totalul valorilor seriei. Nu trebuie totuși să omitem faptul că ea se referă la abaterea datelor față de mediană și nu ne spune nimic despre abatera datelor față de media aritmetică (se poate întâmpla ca valoarea mediei aritmetice să nu se afle între prima și a treia cuartilă).

Sintetizarea variației unei caracteristici într-o singură expresie numerică se realizează prin valoarea

Scanned with OKEN Scanner

medie calculată pe baza abaterilor individuale ale variantelor de la media lor.

IV.3. Abaterea medie(A_m)

Abaterea medie (absolută) se calculează ca o medie aritmetică simplă sau ponderată a tuturor abaterilor termenilor seriei de la media lor. Relațiile de calcul sunt:

$$A_m = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n},$$

pentru o serie simplă,

și

$$A_m = \frac{\sum_{i=1}^{n} |x_i - \bar{x}| f_i}{\sum_{i=1}^{n} f_i}$$

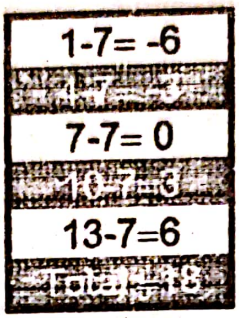
pentru o serie cu frecvențe.

Să considerăm următoarea serie:

						cu media 7
II.	1	4	7	10	13	tot cu media 7.

Scanned with OKEN Scanner

Pentru primul șir abaterea mediei este zero (7-7=0). Pentru cel de-al doilea avem:



Deci, A_m= 18/5 = 3,6 și se exprimă în unitățile de măsură ale caracteristicii.

Se remarcă faptul că diferențele $x_i - \bar{x}$ pot lua valori pozitive (când valorile variabilei sunt mai mici decât media aritmetică, $x_i < \bar{x}$) sau valori negative (în caz contrar, respectiv $x_i > \bar{x}$). De aceea diferențele $x_i - \bar{x}$ se iau în valoare absolută (în modul) fapt care împiedică prelucrările statistice ulterioare. Acest aspect determină folosirea celorlalți indici statistici pentru caracterizarea dispersiei.

IV.4. Dispersia (variația, σ^2)

Cea mai bună și mai des folosită măsură a variabilității este dispersia sau rădăcina ei pătrată, abaterea medie pătratică sau abaterea standard.

Ca și media aritmetică, dispersia se notează cu două simboluri:

a) σ^2, reprezentând dispersia unei populații;

b) S^2, dispersia unui eșantion.

Dispersia se calculează ca medie aritmetică a pătratelor abaterilor termenilor față de media lor:

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}$$

pentru date negrupate,

și

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 f_i}{\sum_{i=1}^{N} f_i}$$

pentru o serie cu frecvențe.

Determinarea dispersiei presupune, după cum rezultă din formule, efectuarea următoarelor calcule:

- stabilirea abaterilor individuale ($x_i - \overline{x}$);
- ridicarea la pătrat a abaterilor individuale (dispar astfel valorile negative);

Scanned with OKEN Scanner

- înmulțirea pătratelor abaterilor cu frecvențele de apriție (numai în cazul dispersiei calculată pentru o serie cu frecvențe);
- însumarea pătratelor abaterilor (multiplicate cu frecvențele);
- raportarea totalului la numărul total de unități (cazuri) înregistrate.

Considerând seria II de la punctul anterior (IV.3), să exemplificăm ($\bar{x} = 7$):

	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
	-6	36
	-3	9
	0	0
	3	9
	6	36
Total	0	90

Dacă șirul ar fi fost mult mai mare atunci împărțirea s-ar fi făcut la N. În cazul nostru, al unui eșantion, corect este:

$$S^2 = \frac{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}{n-1}; S = \sqrt{S^2}$$

Expresia *n-1* este numită "**grade de libertate**". Ea semnifică numărul vaorilor independente care sînt

necesare pentru determinarea divizorului. După cum știm (vezi proprietățile mediei aritmetice, cap.III.1.3.) suma abaterilor de la media aritmetică este întotdeauna egală cu zero. Presupunând că, din cele cinci observații de mai înainte, le cunoaștem doar pe primele patru (-6, -3, 0, 3), cea de-a cincea abatere de la media aritmetică trebuie să fie neapărat 6 pentru ca suma abaterilor de la media aritmetică să fie egală cu zero. Rezultă că numai patru din cele cinci observații sînt independente și deci le putem considera ca fiind grade de libertate, pe când cea de-a cincea observație este în mod necesar determinată de valorile primelor patru. Dacă am mai avea încă un eșantion cu n=1000 de valori, împărțirea sumei pătratelor abaterilor individuale la 1000 sau la 999 (1000-1) va da cam același rezultat, pe când împărțirea prin 5 (a unei sume de 90, de exemplu) sau prin 4 va da rezultate foarte diferite (18 în loc de 22,5). Pentru a simplifica lucrurile, în toate cazurile se folosește n-1 indiferent de mărimea eșantionului.

IV.5. Abaterea standard (abaterea medie pătratică)

Abaterea standard (medie patratică, deviație standard, abatere tip) se calculează ca o medie pătratică

Scanned with OKEN Scanner

din abaterile valorilor individuale de la media lor aritmetică sau, cu alte cuvinte, extrăgând rădăcina patrată din dispersie (σ^2).

Formulele de calcul sînt:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}}$$

și

$$S = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n-1}}$$

pentru datele negrupate referitoare la întreaga populație, respectiv pentru un eșantion (S)

și

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2 f_i}{\sum_{i=1}^{N} f_i}},$$

pentru o serie cu frecvențe.

Abaterea medie pătratică se exprimă în unitățile de măsură ale caracteristicii și este mai mare decât cea medie deoarece prin ridicare la pătrat se acordă o importanță mai mare abaterilor mai mari. Ca urmare, cele

Scanned with OKEN Scanner

două abateri (medie, standard) nu pot fi folosite pentru compararea gradului de variație a unor caracteristici exprimate în unități de măsură diferite. În acest caz este necesar să se calculeze un indicator deosebit de prețios - coeficientul de variabilitate (de variație) - care reprezintă exprimarea procentuală a câtului dintre abaterea standard și media aritmetică:

$$V = \frac{S}{\bar{x}} \times 100$$

Principala calitate a coeficientului de variabilitate constă în aceea că face posibilă compararea colectivelor măsurate cu unități de măsură diferite (secunde cu metri, de exemplu) sau a diferitelor colective măsurate cu aceeași probă (sau probe diferite).

Cu cât coeficientul de variație se apropie mai mult de zero, cu atât variația este mai mică și colectivitatea mai omogenă, respectiv media mai reprezentativă. În schimb, cu cât coeficientul de variabilitate este mai mare cu atât variația este mai intensă, iar gradul de semnificație al mediei este mai scăzut.

Scanned with OKEN Scanner

Statistica recomandă pentru interpretare următoarele trepte de apreciere:

- 0 - 10%, omogenitate mare;
- 10 -20%, omogenitate medie;
- peste 20%, lipsă de omogenitate.

Agoston (citat de Șt. Tudös, p.33) propune (pentru procesul instructiv-educativ) următoarea scară de variație:

- 0 - 15%, împrăștiere mică;
- 15 - 25%, împrăștiere moderată;
- 25 -35%, împrăștiere mare;
- peste 35%, împrăștiere excesivă.

Aceste valori sunt influențate de numărul de cazuri și de amplitudine, găsind o variație mai mare la un număr mare de cazuri și la probele ce permit obținerea de rezulate mult diferite.

Exercițiu

Să se calculeze abaterea standard S pentru următoarele șiruri de date:

1.	8, 15, 13, 6, 10, 16, 7, 12, 11, [illegible]
2.	12, 10, 18, 13, 7, 8, 11, 15, 8, [illegible]
3.	9, 8, 9, 15, 8, 9, 11, 9, 13, [illegible]
4.	12, 28, 19, 15, 15, 26, 18, [illegible]
5.	7, 18, 20, 14, 27, 23, 13, [illegible]

Răspuns:

	1.	2.	3.	4.	5.
S	±3,31	±4,64	±3,36	±8,09	±8,03

CAPITOLUL V

DISTRIBUȚIA NORMALĂ

V.1.Distribuția (curba) normală și caracteristicile sale

Unul din scopurile prelucrării datelor de observații sub formă de distribuții de frecvențe este acela de a face comparații între distribuția (și curba aferentă) empirică și una teoretică, fixată ca model de referință. Similitudinea de formă a celor două tipuri de curbe (empirică și teoretică) este foarte importantă, căci ea oferă niște posibilități de analiză statistică complexă și aprofundată în urma căreia se pot trage multe concluzii de ordin științific. La originea multor astfel de analize se află tocmai curba normală, cu proprietățile sale.

Principiul curbei normale se bazează pe probabilitatea de apariție a unui eveniment atunci când acesta depinde de șansă. De exemplu, când arbitrul "dă cu banul" pentru alegerea terenului de joc, șansa de a obține stema (să zicem) este una din două, aceeași cu cea de a "cădea" banul (moneda nu are decât două fețe, nu-i așa?). Mergând mai departe, dacă se aruncă două

Scanned with OKEN Scanner

monede (sau o aceeași monedă de două ori) sunt patru posibilități, după cum urmează:

1		2		3		4	
a	b	a	b	a	b	a	b
S	S	S	B	B	S	B	B

unde S=stema, B=banul

Așadar, șansa de a ieși stema de două ori este una din patru (0,25 sau 25%) - la fel și pentru ban; de a cădea stema sau banul este una din două (0,5 sau 50%). Continuând raționamentul vom vedea că este doar o șansă din opt (2^3) să cadă stema de trei ori la rând atunci când aruncăm trei monede, și o șansă din 1024 (2^{10}) când aruncăm 10 monede. Dacă am plasa punctele corespuzătoare șanselor de a obține stema sau banul (la un număr cât mai mare de aruncări ale monezii) pe un sistem de axe ortogonale și le-am uni apoi între ele ar rezulta curba normală (în care șansele de obține numai stema sau numai banul ar fi plasate la extremități și stema sau banul spre centru) definită de relația matematică proprie (pe care nu o prezentăm).

Această teorie a distribuției normale, aplicată la șansa de apariție a stemei sau banului are aplicabilitate și în domeniul socio-uman. Erediatea, mediul bio-psiho-social, starea de antrenament, sunt factori de care depind

Scanned with OKEN Scanner

numeroase atribute umane, șansa jucând de foarte multe ori un rol deosebit de important.

Grafic, distribuția normală se prezintă ca o curbă normală. **Curba normală** reprezintă de fapt o distribuție de frecvență în care cele mai multe frecvențe sunt concentrate în centrul ei, iar pe măsură ce se depărtează de centru devin, simetric, tot mai puține, apropiindu-se asimptotic de zero, ceea ce înseamnă că ea nu atinge niciodată axa orizontală sau abscisa. Curba normală este simetrică și are forma unui clopot (*clopotul lui Gauss*). Ea este împărțită în două părți egale de media aritmetică a frecvențelor pe care le reprezintă grafic; aceasta înseamnă că media aritmetică a curbei normale este egală cu zero (media aritmetică devine originea sistemului de axe).

În cazul curbei normale, media, mediana și modulul sunt identice. Dat fiind că suma frecvențelor este egală cu 1, rezultă că și suprafața totală aflată sub curba normală este egală cu 1 (există tabele ale suprafeței curbei normale care arată valoarea proporțiilor acestor suprafețe - vezi tabel 5.1). În figura 5.1. prezentăm o curbă normală cu o medie aritmetică egală cu zero și suprafața de sub ea egală cu unitatea, având două puncte de inflexiune:

Scanned with OKEN Scanner

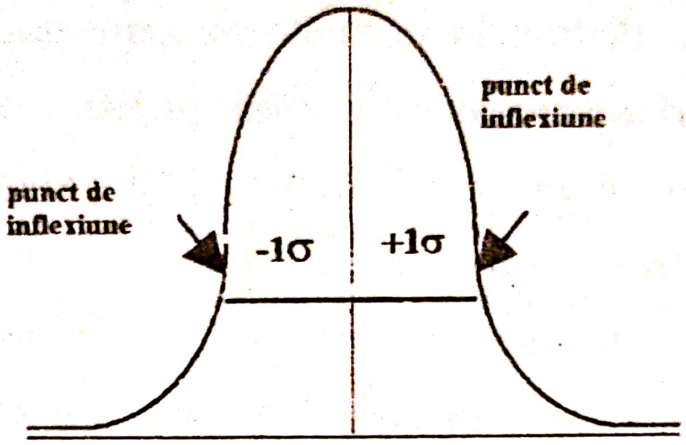


Figura 5.1 - Curba normală cu cele două puncte de inflexiune

Se disting două puncte de inflexiune, simetric amplasate, unul în dreapta și celălalt în stânga mediei. În matematică punctul de inflexiune este definit ca punctul care își schimbă alura de la convexitate la concavitate și invers. Dacă o distribuție poate fi privită ca normală, atunci distanța dintre medie și punctul de inflexiune din dreapta mediei acoperă 0,3413 din suprafața totală de sub curba normală și este egală cu +1σ după cum se vede în figura de mai sus. Curba normală fiind simetrică, rezultă că 0,3413 din suprafața ei situată la stânga mediei aritmetice este egală cu -1σ. Deci, o proporție de 68,26% din suprafața totală de sub curba normală este inclusă la o distanță de plus sau minus o abatere standard față de

media aritmetică (sau, cu alte cuvinte, aproximativ două treimi din suprafața totală este inclusă în acest interval de $\pm 1\sigma$). Aceasta este distanța dintre cele două puncte de inflexiune.

Dacă măsurăm două abateri standard de o parte și de alta a mediei, atunci se includ 95,46% din suprafața de sub curba normală, iar dacă măsurăm trei abateri standard de fiecare parte a mediei, includem 99,72 procente din suprafata totală. Suprafețele de sub curba normală se găsesc în tabele special întocmite pentru o utilizare rapidă Vezi (tabelul 5.1 și figura 5.2).

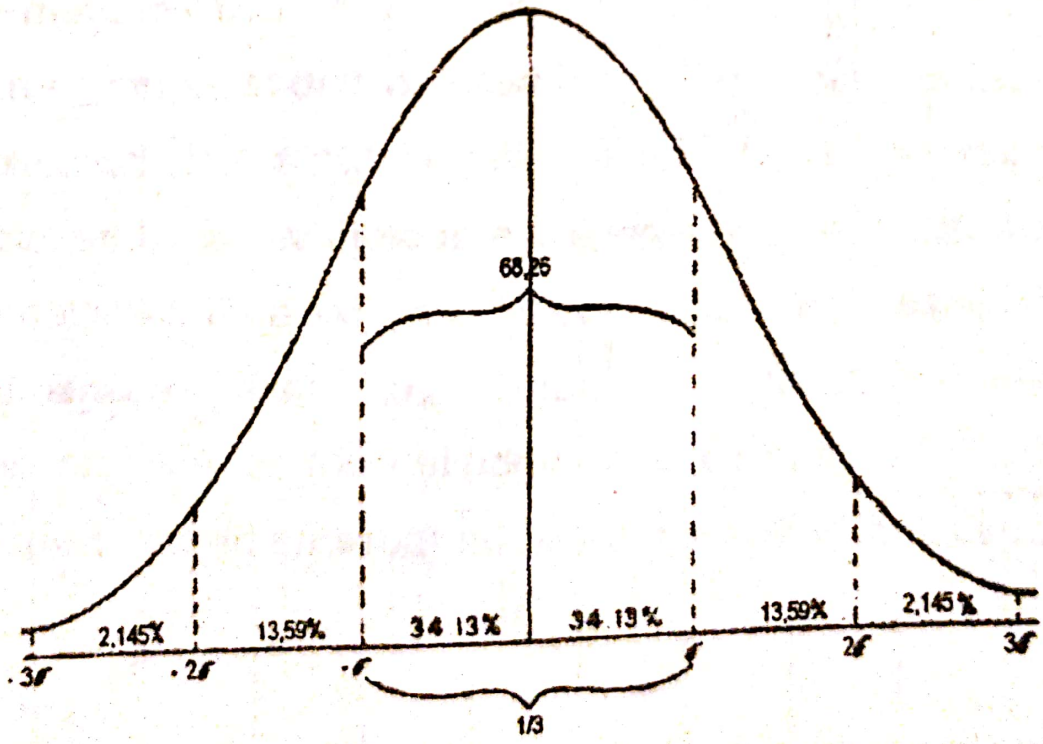


Figura 5.2 Curba normală cu unități de abateri standard

De exemplu, dacă vom căuta pe coloana z valoarea **1,0** (deci 1σ) vom găsi pe coloana următoare **(,00)** valoarea 3413, adică 34,13% din suprafața de sub curba normală luată la dreapta mediei; la fel, pentru 2σ vom găsi la 2,0 pe coloana **(,00)** alăturată valoarea 4773 (47,73% x 2 =95,46%), iar pentru 3σ - 49,86%, respectiv 99,72% din suprafața de sub curbă luată de o parte și de alta a mediei aritmetice (49,86% x 2).

Dacă distribuția empirică aproximează o distribuție normală, abaterea standard devine un instrument de măsură deosebit de important (vezi capitolul VI).

Desigur, multe din distribuțiile cu care avem de-a face în realitate, nu aproximează o distribuție normală. Dar însemnătatea fundamentală a distribuției normale în statistică și, în consecință și a abaterii standard, rezultă din faptul că chiar dacă populațiile nu prezintă caracteristicile unei distribuții normale, variabilele eșantioanelor respective tind să fie distribuite normal. Aceasta este o lege cunoscută sub numele de **teorema limitei centrale.**

Scanned with OKEN Scanner

Suprafețe sub curba normală

Tabel 5.1

	0000	0040	0080	0120	0159	0199	0239	0279	0319	0359
	0398	0438	0178	0517	0557	0596	0636	0675	0714	0753
	0793	0832	0871	0910	0948	0987	1026	1064	1103	1141
	1179	1217	1255	1293	1331	1368	1406	1443	1480	1517
	1554	1591	1628	1664	1700	1736	1772	1808	1844	1879
	1915	1950	1985	2019	2054	2088	2123	2157	2190	2224
	2257	2291	2324	2357	2389	2422	2454	2486	2518	2549
	2580	2612	2642	2673	2704	2734	2764	2794	2823	2852
	2881	2910	2939	2967	2995	3023	3051	3078	3106	3133
	3159	3186	3212	3238	3264	3289	3315	3340	3365	3389
	3413	3438	3461	3485	3506	3531	3554	3577	3599	3621
	3643	3665	3686	3718	3729	3749	3770	3790	3810	3830
	3849	3869	3886	3907	3925	3944	3962	3980	3997	4015
	4032	4049	4066	4083	4099	4115	4131	4147	4162	4177
	4192	4207	4222	4236	4251	4265	4279	4292	4306	4319
	4332	4345	4357	4370	4382	4394	4406	4418	4430	4441
	4452	4463	4474	4485	4495	4505	4515	4525	4535	4545
	4554	4564	4573	4582	4591	4599	4608	4616	4625	4633
	4641	4649	4656	4664	4671	4678	4686	4693	4699	4706
	4713	4719	4726	4732	4738	4744	4750	4758	4762	4767
	4773	4778	4783	4788	4793	4798	4803	4808	4812	4817
	4821	4826	4830	4834	4838	4842	4846	4850	4854	4857
	4861	4865	4808	4871	4875	4878	4881	4884	4887	4890
	4893	4896	4898	4901	4904	4906	4909	4911	4913	4916
	4918	4920	4922	4925	4927	4929	4931	4932	4934	4936
	4938	4940	4941	4943	4945	4946	4948	4949	4951	4952
	4953	4955	4956	4957	4959	4960	4961	4962	4963	4964
	4965	4966	4967	4968	4969	4970	4971	4972	4973	4974
	4974	4975	4976	4977	4977	4978	4979	4980	4980	4981
	4981	4982	4983	4984	4984	4984	4985	4985	4986	4986
	4986	4987	4987	4988	4988	4988	4989	4989	4989	4990
	4990	4991	4991	4991	4992	4992	4992	4992	4993	4993
	4993									
	4995									
	4996									
	4997									

Scanned with OKEN Scanner

V.2. Parametrii formei

Acești parametrii se referă în primul rând la forma curbelor (poligoanelor) frecvențelor. Media aritmetică, mediana și modulul sunt identice numai în cazul unei distribuții perfect simetrice. Pe măsură ce o distribuție devine mai asimetrică valorile mediei aritmetice, a medianei și a modulului tind să fie divergente.

V.2.1. Indici de asimetrie

Indicii de asimetrie exprimă numeric măsura în care maximul de frecvență al unei distribuții este deplasat spre stânga, sau spre dreapta, față de centrul intervalului de variație ai caracteristicii.

În cazul unei distribuții moderat asimetrice există o relație aproximativă între mediană, modul și medie. Într-un asemenea caz, mediana se află la o treime a distanței între media aritmetică și modul, sau altfel exprimat:

$$Mo = \bar{x} - 3(\bar{x} - Me).$$

Scanned with OKEN Scanner

Deoarece media aritmetică tinde să fie împinsă în direcția valorilor extreme ale "cozii" celei mai lungi a unei distribuții, asimetria poate fi măsurată prin mai mulți coeficienți din a căror multitudine prezentăm aici doar doi:

1) *coeficientul lui Pearson*, se calculează pentru distribuțiile de frecvențe când ia în considerare modulul:

$$C_{as} = \frac{\left(\bar{x} - M_o\right)}{\sigma}$$

În funcție de raporturile de mărime dintre medie și modul coeficientul poate fi:

a) negativ (asimetrie negativă sau de dreapta) - $\bar{x} < Mo$ - (numărătorul din formulă este mai mic ca zero), ceea ce înseamnă că maximul de frecvență este deplasat spre dreapta față de centrul intervalului de variație;

b) nul (asimetrie nulă), curba de distribuție este perfect simetrică;

c) pozitiv (asimetrie pozitivă sau de stânga) - $\bar{x} > Mo$ - modulul este plasat la stânga față de medie și față de centrul intevalului în cadrul intervalului de variație.

2) *coeficientul lui Fisher* (γ_1) care ia în considerare cubul abaterilor valorilor de la media lor aritmetică și a

cărui interpretare, în ceea ce privește asimetria, se realizează la fel că în cazul coeficientului lui Pearson:

$$\gamma_1 = \frac{\sum\left(x_i - \bar{x}\right)^3}{n\sigma^3},$$

pentru șiruri de valori individuale și

$$\gamma_1 = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3 f_i}{\sigma^3 \sum_{i=1}^{n} f_i}$$

pentru distribuția pe clase de frecvențe

V.2.2 Indicii de aplatizare (de exces)

Acești indici (coeficienți) exprimă numeric gradul de aplatizare a curbei de distribuție, aplatizare ce depinde de gradul de concentrare a valorilor caracteristicii în apropierea maximului de frecvență; cu alte cuvinte, această aplatizare depinde de diferența dintre frecvența clasei modale și cea a claselor din apropiere (cu cât diferența este mai netă, cu atât curba va fi mai ascuțită, și invers).

Scanned with OKEN Scanner

Pentru măsurarea asimetriei de aplatizare (de exces) cel mai des utilizat este coeficientul γ_2 al lui Fisher:

$$\gamma_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n\sigma^4} - 3$$

pentru șiruri de valori individuale și

$$\gamma_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4 f_i}{\sigma^4 \sum_{i=1}^{n} f_i} - 3$$

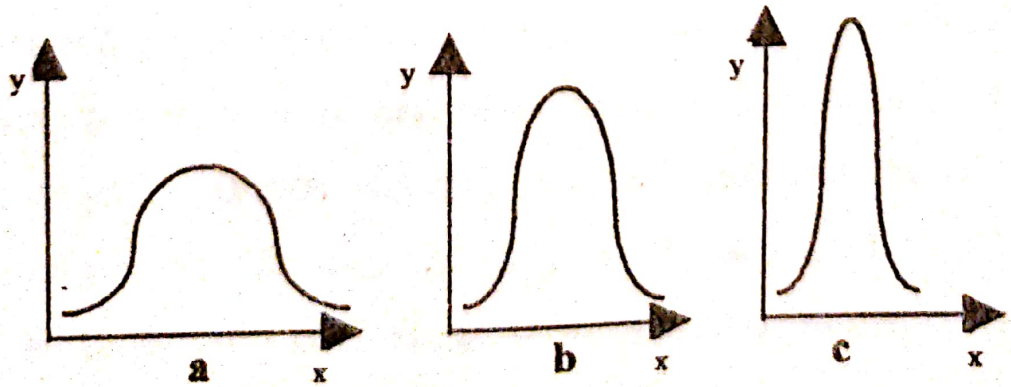
pentru distribuția pe clase de frecvențe.

Acest indice poate fi:

a) <u>negativ</u> - $\gamma_2 < 0$ - curba fiind mai aplatizată decât una normală (sau *platikurtică*);

b) <u>nul</u> - $\gamma_2 = 0$ - curba este una de tip normal.

c) <u>pozitiv</u> - $\gamma_2 > 0$ - curba prezintă o boltire excesivă, mai ascuțită decât cea normală (*leptokurtică*).



Scanned with OKEN Scanner

Exerciții:

1. Se aruncă o monedă de patru ori (sau patru monede în același timp). Care este probabilitatea de a obține:
   a) secvența B S B S?;
   b) de patru ori stema?;
   c) de trei ori banul ?

Răspunsuri

1. a - 1/16;
   b - 1/16;
   c - 1/4.

Scanned with OKEN Scanner

CAPITOLUL VI

VARIABILELE NORMATE

(ETALONAREA)

Etalonarea este acțiunea de transformare a datelor inițiale în așa fel încât să devină posibile comparațiile între două sau mai multe distribuții empirice, sau între o distribuție empirică și una teoretică, idiferent de ordinul de mărime, dispersie sau unități de măsură. Adesea, în investigarea aceluiași fenomen, suntem puși în situația de a compara rezultatele individuale din două distribuții de valori ale variabilei obținute prin măsurări succesive ale aceluiași colectiv sau de a compara două colective diferite prin rezultatele înregistrate în urma unei singure măsurări. În urma recoltării datelor, profesorul (antrenorul) alcătuiește un clasament la fiecare probă și constată că un subiect se clasează la o probă pe locul 7, la altă probă pe locul 5 și la alta pe primul loc. Această situație se aplică și la celelalte unități ale colectivului rezultând faptul că este necesar ca fiecare subiect să fie apreciat în contextul colectivului din care face parte prin prisma performanțelor

Scanned with OKEN Scanner

realizate. Făcând suma locurilor din clasament (5+7+1) se obține un clasament al membrilor colectivului testat cu mai multe probe, dar se pierde aspectul calitativ, dat de obținerea unei performanțe de excepție la una dintre probe. Așa încât este necesar să identificăm modalități de stabilire a unor criterii care să surprindă atât latura cantitativă cât și pe cea calitativă.

Metodele statistice intervin în rezolvarea acestei probleme prin procedeele de transformare a valorilor brute ale variabilei în variabile normate ce reprezintă unități ce pot exprima poziția unui subiect într-o distribuție dată, atât în raport cu valorile centrale (media aritmetică) cât și cu cele ale dispersiei (abaterea standard).

Se cunosc mai multe tipuri de variabile normate, noi prezentându-le doar pe cele mai utilizate, respectiv variabilele Z, T și H.

VI.1. Variabilele Z

Pentru a exprima mărimile unei distribuții luate în considerare prin măsuri Z, se împarte diferența dintre

Scanned with OKEN Scanner

oricare din acele mărimi și media aritmetică a distribuției prin abaterea standard. În cazul unei populații avem:

$$z = \frac{X - \overline{X}}{\sigma},$$

iar în cazul unui eșantion

$$z = \frac{x - \overline{x}}{S}.$$

În scările variabilelor standard Z, media aritmetică se reduce la valoarea zero, iar abaterea standard la valoarea 1 (σ=1).

Să presupunem că un student a obținut la un colocviu de Teoria educației fizice și sportului nota 7, la care întreaga grupă a obținut în medie nota 5 cu o abatere standard egală cu 1 și la alt colocviu, de Metodica educației fizice, nota 9, colegii săi obținând în medie nota 6, cu o abatere standard de 2 (considerând că variabilele sunt distribuite normal, putem spune că 68,26% dintre aceștia au obținut note între 4 și 8, $\sigma\pm2$). La prima vedere sîntem tentați să afirmăm că studentul nostru a obținut o notă mai bună la Metodică decât la Teoria educației fizice; dacă măsurăm însă aceste performanțe cu măsuri

Scanned with OKEN Scanner

standard, nota lui la al cel de-al doilea colocviu este de 1,5 măsuri standard peste medie, în timp ce la primul este de două măsuri standard peste medie:

Colocviu de Teoria educației fizice -

$$z = \frac{7-5}{1} = 2;$$

Colocviu de Metodica educației fizice -

$$z = \frac{9-6}{2} = \frac{3}{2} = 1,5$$

Sau, alt exemplu (Dragnea A., p.33), o variabilă care exprimă numărul de centimetri realizați la săritura în lungime de pe loc de un subiect care a realizat performanța de 224 cm, ceilalți membrii ai colectivului sărind în medie 217 cm cu o abatere standard de 12 cm. Vom avea:

$$z = \frac{224-217}{12} = \frac{7}{12} = 0,58$$

Valoarea obținută fiind pozitivă, rezultă că ea se situează în dreapta mediei, într-o abatere standard spre dreapta. Revenind la tabelul 5.1 (vezi capitolul anterior) valorii pozitive 0,58 îi corespunde un procent de 21,9% din cazuri. Dacă luăm altă performanță, 204 cm de exemplu. rezultă:

Scanned with OKEN Scanner

$$z = \frac{204 - 217}{12} = -\frac{13}{12} = -1{,}08$$

deci o valoare negativă, situată în stânga mediei pe curba lui Gauss.

Variabilele normate pe scara Z sunt deseori incomode prin apariția semnului "-" (minus) (când media este mai mare decât valoarea individuală luată în calcul, $\bar{x}$>x) și a valorilor zecimale. Aceste neajunsuri sunt eliminate prin utilizarea variabilelor normate T (introdusă de McCall) și H (propusă de Hull).

VI.2. Variabilele T

Scara propusă de McCall are o distribuție cu media $\bar{x}$=50 și σ=10, iar variabilele T se obțin utilizând relația:

$$T = 50 + \frac{10}{\sigma}\left(x - \bar{x}\right)$$

Propriu-zis, pe această scară fiecare unitate T este egală cu o zecime dintr-o unitate σ.

Scanned with OKEN Scanner

VI.3. Variabilele H

Datorită faptului că o distribuție depășește foarte rar intervalul ±3,5σ, Hull a introdus o altă scară de variabile normate care apelează tot la media $\bar{x}$ =50, dar σ=14, relația de calcul a variabilelor H fiind:

$$H = 50 + \frac{14}{\sigma}\left(x - \bar{x}\right)$$

Folosindu-ne de exemplele de la variabilele Z (săritura în lungime, cu performanțele X_1=224 cm, X_2=204 cm, media 217 cm și abaterea standard 12 cm), să facem o comparație între cele trei scări:

	Z		H
	0,58		58,16
	-1,08		34,84

Se remarcă faptul că este destul de dificil să caracterizezi diferența între performanțele stadardizate ale celor doi subiecți (X_1 și X_2) recurgând la variabila Z, fiind mai ușoară comparația utilizând scările introduse de McCall și Hull.

Utilizând una din scările prezentate, aceasta se aplică la toate probele la care sunt supuși subiecții, sumându-se apoi punctele realizate. În acest fel este mai

ușor de alcătuit un clasament, deoarece avem situația fiecărui subiect în contextul respectiv sau, cu alte cuvinte, măsurăm toate performanțele cu un etalon unic.

Exercițiu:

Să se calculeze scorurile standard pentru următorul șir de date, reprezentând notele obținute de o clasă de elevi la trecerea unei probe de control:

Nota
3
4
5
6
7
8
9
10

Răspuns:

Nota	Z	H
3	-1.696	26.253
4	-1.124	34.252
5	-0.553	42.251
6	0.0178	50.249
7	0.589	58.248
8	1.160	66.247
9	1.731	74.246
10	2.303	82.244

Scanned with OKEN Scanner

CAPITOLUL VII

CORELAȚIA STATISTICĂ

De multe ori, specialiștii din domeniul educației fizice și al antrenamentului sportiv se întreabă în ce măsură exercițiile preconizate sau folosite în pregătirea sportivilor influențează randamentul acestora în probele de concurs; și invers, cei care obțin rezultate bune în probele de concurs obțin, în aceeași măsură, rezultate și la exercițiile pregătitoare?

Pentru rezolvarea unor asemenea probleme este necesară cunoașterea măsurii, a gradului de dependență a fenomenelor, mai exact a posibilității în care existența unei variabile o determină pe cealaltă. Metodele statistice intervin în analiza cantitativă și calitativă a legăturilor dintre fenomene prin tehnicile de corelație. În cercetările curente ne interesează în special relații de tipul:

a) cauză - efect, când variația fenomenului **A** determină variația fenomenului **B**, între ele stabilindu-se legături funcționale;

Scanned with OKEN Scanner

b) paralel independente, când fenomenele **A** și **B** sînt condiționate de fenomenul **C**, între primele două variația fiind independentă;

c) relații independente, când fenomenul **A** variază independent de fenomenul **B**.

La toate acestea găsim rezolvarea prin înțelegerea sensului corelației și prin calcularea coeficienților de corelație.

VII.1. Coeficientul de corelație și de determinare

VII.1.1 Coeficientul de corelație

Prin **corelație** se înțelege procedeul statistic prin care se stabilește raportul de dependență dintre două fenomene.

În funcție de *direcția de variație* a fenomenelor, corelația poate fi pozitivă dacă variația este concomitentă directă (creșterea variabilei determinante x duce la creșterea variabilei determinate y) sau negativă, dacă variația este concomitentă indirectă (creștera variabilei x

Scanned with OKEN Scanner

determină descreșterea variabilei y). În cercetarea relației dintre fenomene este necesar să se stabilească *gradul de corelație*, respectiv măsura în care creșterea variabilei x determină creștera sau descreșterea variabilei y, *sensul corelației* (pozitivă, negativă sau nulă) și *natura* sa, respectiv dacă fenomenele sînt în relație cauzală sau în relații de independență. ambele fiind condiționată de un al treilea fenomen. Gradul și sensul corelației se stabilesc prin intermediul coeficienților de corelație, iar natura corelației rezultă din analiza și interpretarea fenomenelor cercetate. Dar să exemplificăm, pentru a înțelege mai bine afirmațiile de până acum.

Să începem cu un exemplu de corelație (pozitivă) perfectă (reprezentată în figura 7.1), ~~extrem~~ de rar sau de loc întâlnită în practica domeniului nostru și a științelor educației, pornind de la numărul de puncte obținute de 10 subiecți la două teste X și Y:

Subiectul	A	B	C	D	E	F	G	H	I	J
Test Y	4	6	7	8	9	10	11	12	14	15

Se observă ușor că fiecare subiect a obținut la al doilea test (Y) cu 2 puncte mai mult decât la primul test, fără nici o excepție.

Scanned with OKEN Scanner

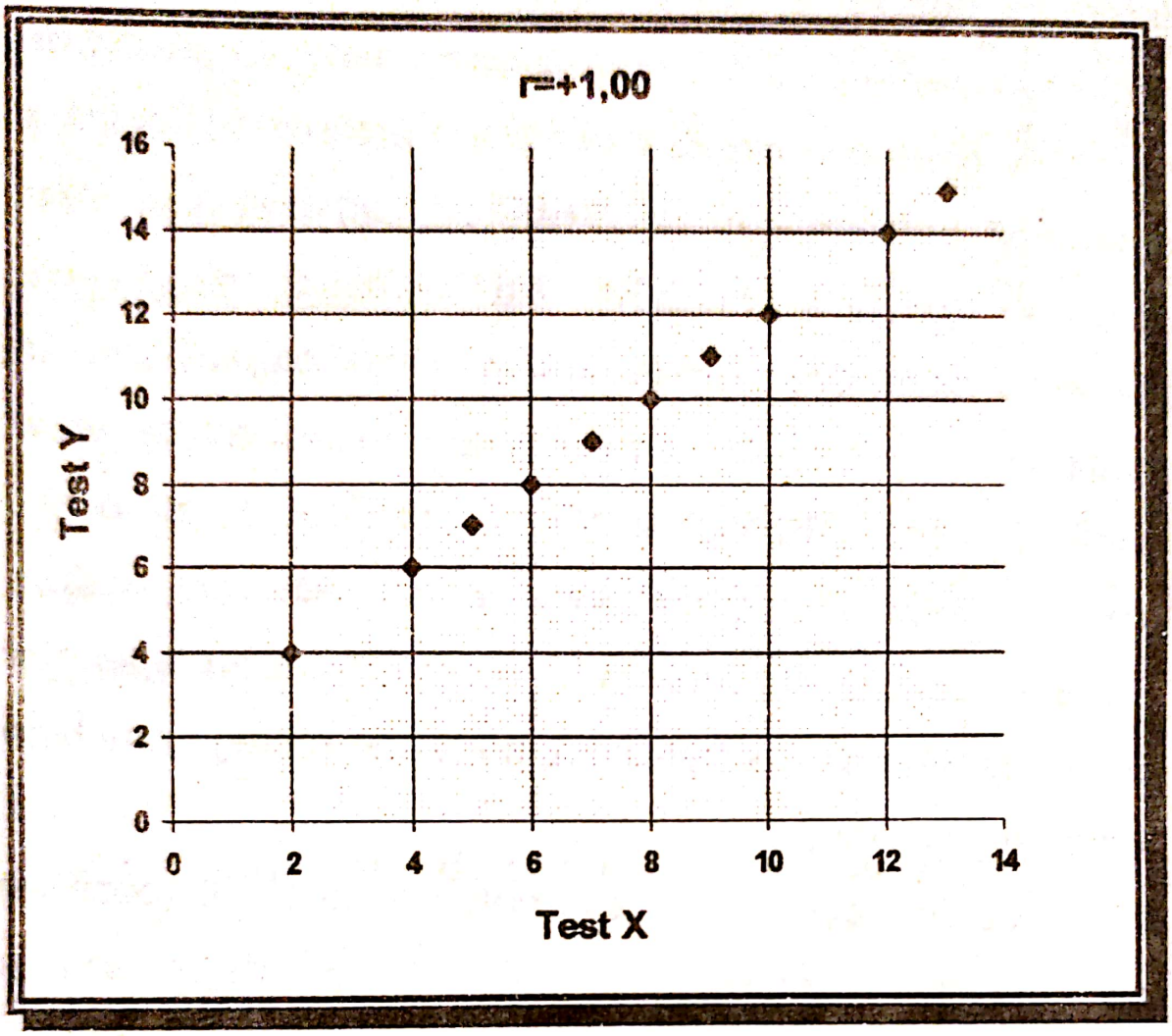


Figura 7.1

O corelație pozitivă puternică (ridicată) se obține atunci când la două teste diferite X și Y, în general, o persoană (subiect) care obține un punctaj bun la testul X obține un scor bun și la cel de-al doilea (Y). Să considerăm următoarele date:

Subiectul	A	B	C	D	E	F	G	H	I	J
Test X	2	3	5	6	8	8	10	10	12	13
Test Y	1	6	7	3	11	9	7	11	14	11

Scanned with OKEN Scanner

și să le reprezentăm grafic în figura 7.2.

Am recurs la reprezentările grafice în cele două figuri pentru a putea remarca diferențele apărute. Astfel, în figura 7.1 punctele sunt coliniare și așezate pe o dreptă ascendentă (spre dreapta, în unghi de 45^0) conform valorilor, pe când în figura 7.2 ele sunt ușor împrăștiate.

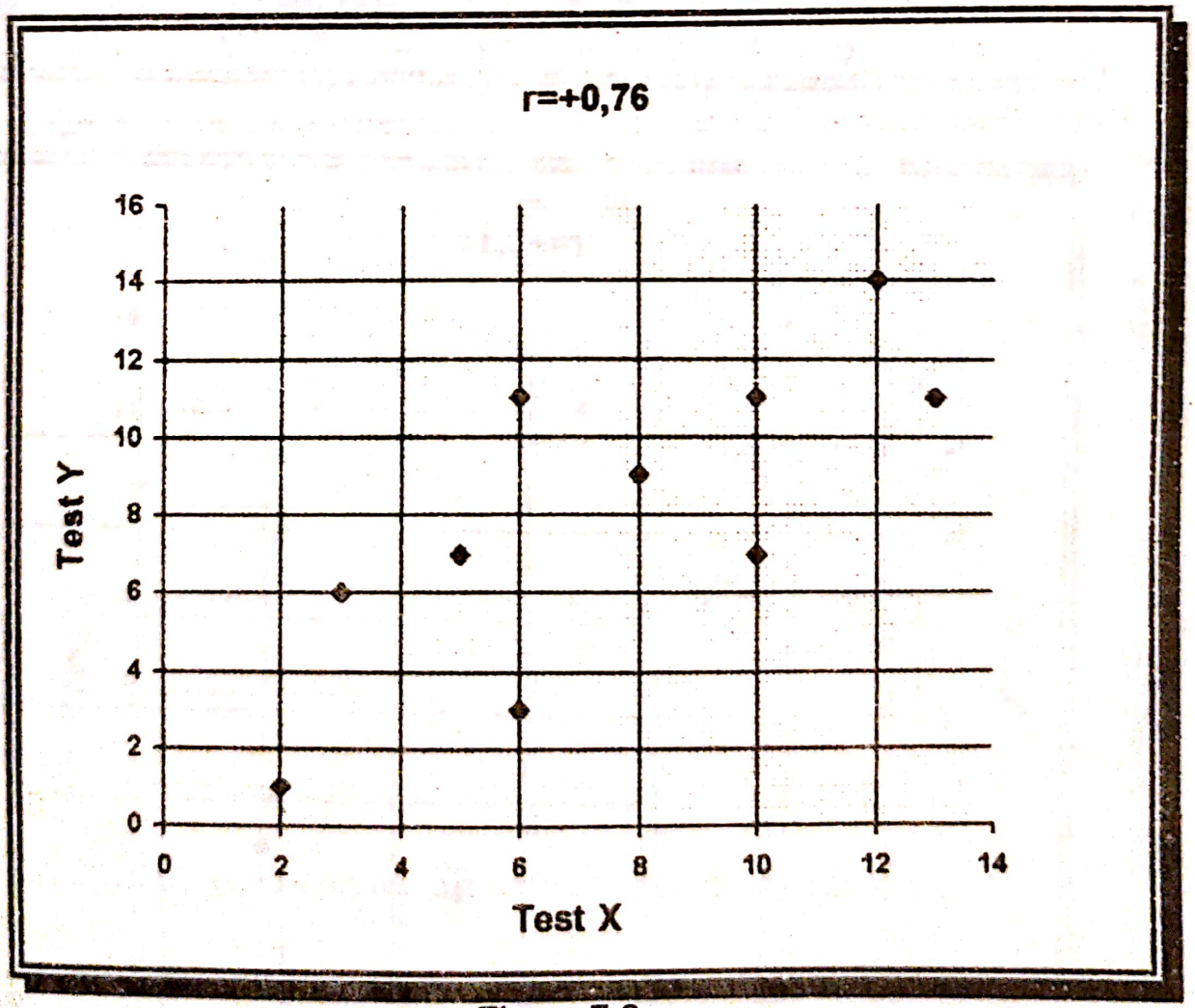


Figura 7.2

Scanned with OKEN Scanner

Un al treilea exemplu îl constituie o corelație pozitivă scăzută (figura 7.3). De această dată un subiect care a obținut un punctaj bun la primul test, a obținut la cel de-al doilea un număr de puncte la fel cu un subiect mai slab cotat inițial. Vom vedea că împrăștierea punctelor este mai mare decât la corelația puternică (pozitivă):

Subiectul	A	B	C	D	E	F	G	H	I	J
Test Y	6	9	2	6	12	2	7	3	11	7

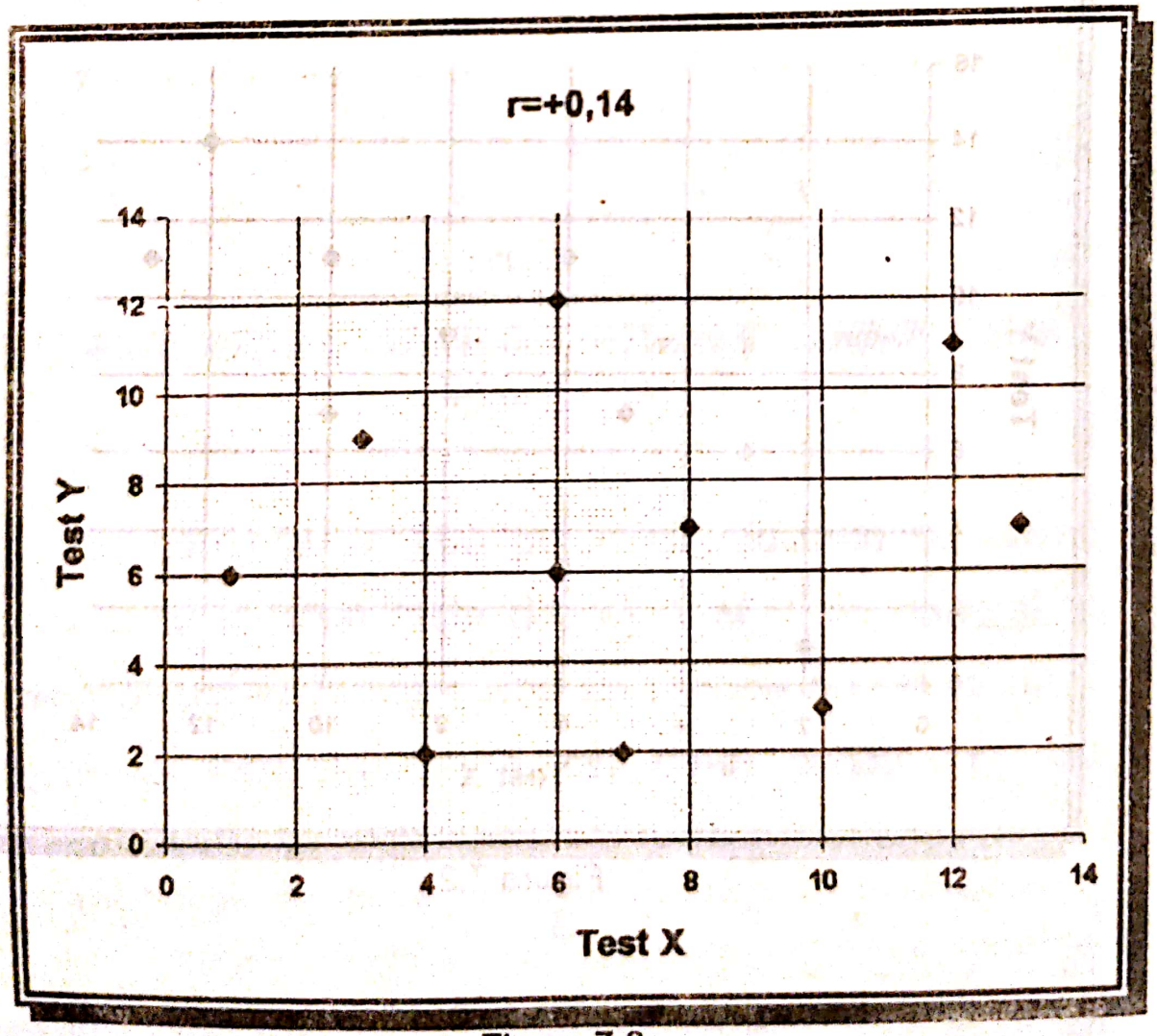


Figura 7.3

Scanned with OKEN Scanner

Ultimul exemplu se referă la cazul unei corelații negative semnificative (remarcabile). Comparând figura 7.4 cu celelalte vom remarca faptul că punctele de pe grafic sunt grupate (dar descendent, spre dreapta), ceea ce arată că un subiect care a obținut un punctaj bun la primul test a obținut rezultate slabe la cel de-al doilea:

Subiecții	A	B	C	D	E	F	G	H	I	J
Test Y	11	13	9	10	12	7	5	8	3	7

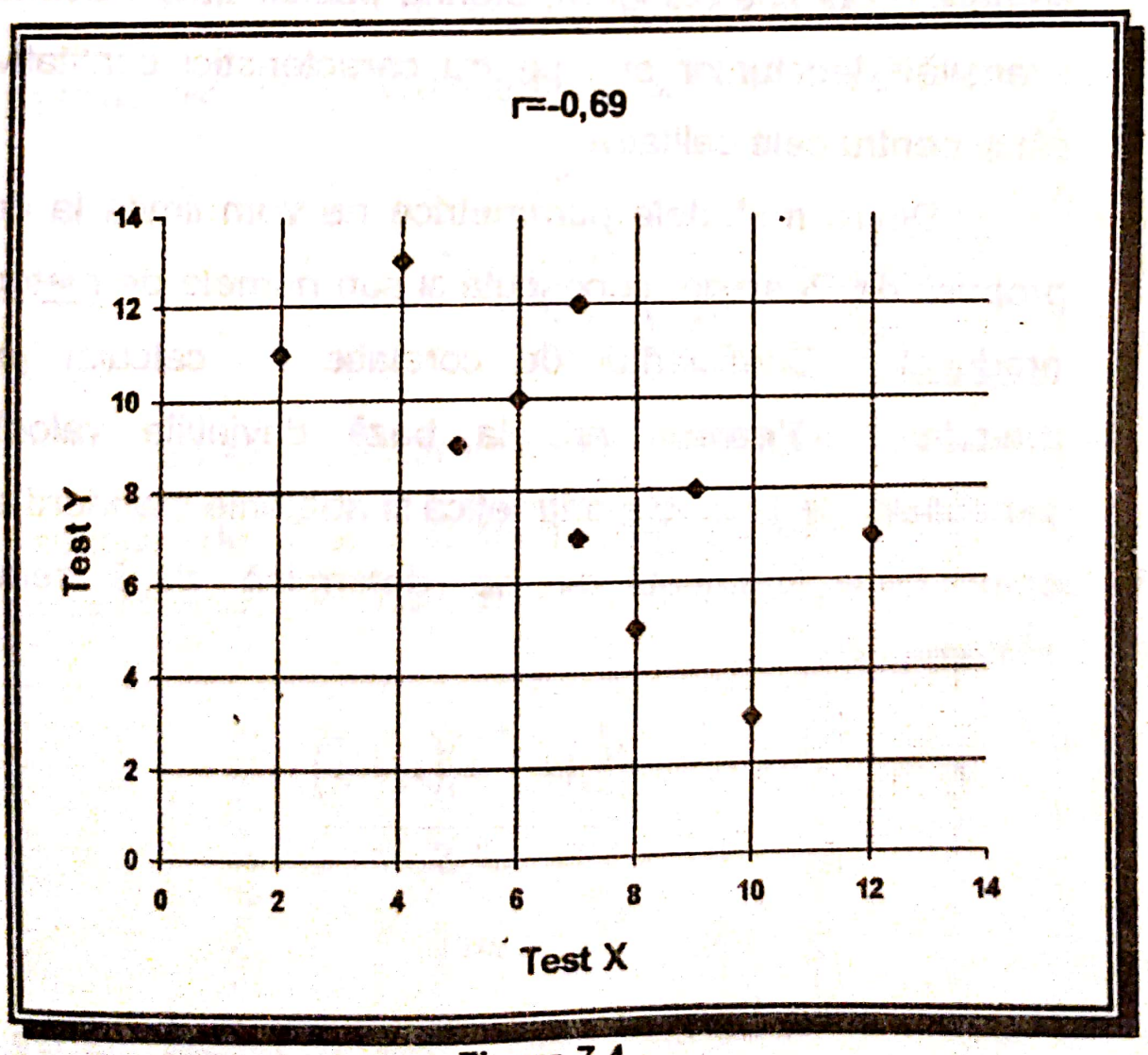


Figura 7.4

Scanned with OKEN Scanner

Gradul și sensul legăturii dintre fennomene se exprimă printr-un indice statistic numit *coeficient de corelație*. Pentru determinarea coeficienților de corelație, statistica matematică pune la dispoziția cercetătorilor diferite tipuri de metode: metode analitice (parametrice) ce țin seama de valorile variabilei și de parametrii curbelor de distribuție (media aritmetică și abaterea standard) și metode neparametrice, care utilizează simboluri sau numere de ordine (ranguri), oferind posibilitatea măsurării intensității legăturilor atât pentru caracteristici cantitative cât și pentru cele calitative.

Dintre metodele parametrice ne vom limita la cea propusă de Pearson, cunoscută și sub numele de <u>metoda produselor</u>. Coeficientul de corelație r calculat prin metoda lui Pearson are la bază deviațiile valorilor variabilelor de la media aritmetică și abaterile standard ale celor două distribuții și se determină după relația matematică:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{nS_x S_y}$$

sau

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

unde - x_i, y_i sînt valorile celor două variabile;
- $\bar{x}$, $\bar{y}$ mediile aritmetice;
- S_x, S_y abaterile standard ale celor două distribuții;
- n reprezentând numărul de cazuri

Notând cu $d_x = x_i - \bar{x}$ și $d_y = y_1 - \bar{y}$ deviațiile valorilor variabilelor de la media aritmetică, formula devine:

$$r = \frac{\sum_{i=1}^{n} d_x d_y}{\sqrt{\sum d_x^2 \sum d_y^2}}$$

Să exemplificăm folosind rezultatele obținute de un eșantion de 10 elevi la două teste X ($\bar{x}$=44) și Y ($\bar{y}$=44):

	33	-11	121		66
	36	-8	64		72
	39	-5	25		30
	42	-2	4		4
	45	1	1		-2
	45	1	1		0
	45	1	1		6
	49	5	25		20
	52	8	64		48
	54	10	100		90
			Σd_x^2=406		$\Sigma d_x d_y$=334

Utilizând ultima formulă obținem $r = \frac{334}{\sqrt{406 \times 330}} = 0{,}912$.

O metodă neparametrică este corelația rangurilor. Ea se bazează pe înlocuirea valorilor empirice ale variabilelor prin rangurile respective, adică prin locul ce-l ocupă în ordinea lor crescătoare sau descrescătoare. Corelația rangurilor poate fi aplicată și pentru calcularea gradului dependenței dintre fenomene și caracteristici calitative care, prin definiție, nu pot fi exprimate numeric, dar care sînt totuși susceptibile de a fi ordonate într-o anumită ierarhie, în cadrul căreia ocupă un anumit rang. Pornim deci de la ipoteza că între rangurile celor două variabile există o dependență a cărei intensitate dorim să o calculăm.

Formula lui Spearman ia în considerație suma pătratelor diferențelor dintre fiecare pereche de ranguri:

$$r = 1 - \frac{6\sum d^2}{n(n+1)(n-1)} = 1 - \frac{6\sum d^2}{n\left(n^2-1\right)}$$

Să luăm de exemplu două șiruri de variabile care reprezintă performanțele obținute de un grup de subiecți la arucarea mingii medicinale de 2 kg cu ambele mâini, înainte și înapoi (Dragnea A., p.40):

Scanned with OKEN Scanner

10	10	10	10	0	0
12	11	8,5	8,5	0	0
14	13	6	5	1	1
9	8	11,5	12	-0,5	0,25
17	15	1	1,5	-0,5	0,25
15	12	4	7.5	-3,5	12,25
					Σ=19,5

Pe baza valorii calculate, putem aprecia că între cele două șiruri de variabile există un înalt grad de legătură deoarece coeficientul este pozitiv, adică 0,931.

Tehnica de lucru, în ordinea operațiilor, este următoarea:

a) se stabilesc rangurile (locurile de ordine) la fiecare șir în parte; în situația în care avem mai multe variabile de aceeași valoare se însumează locurile de ordine și rezultatul se împarte la numărul de variabile cărora le-au fost atribuite (se face media lor aritmetică). De exemplu: la testul X valoarea 15 se repetă de trei ori și se cuvine să acordăm celor trei variabile locurile 3, 4 și

Scanned with OKEN Scanner

5 (mai mare ca 15 sînt 17 și 16) dar nu putem, deoarece performanța este aceeași: 15m. Atunci calculăm media aritmetică 3+4+5=12/3=4 și vom da fiecăruia locul 4; în continuare acordăm locurile corespunzător performanțelor. La fel se procedează și pentru testul Y;

b) se calculează diferențele (d) dintre locurile de ordine (R_x-R_y);

c) se ridică la pătrat (d^2) diferențele obținute, operație prin care dispar valorile negative și se face suma lor.

Înlocuind în formulă se obține:

$$r = 1 - \frac{6 \times 19{,}5}{12(144-1)} = 1 - \frac{117}{1716} = 1 - 0{,}068 = +0{,}931$$

Atenție! La corelație nu se ordonează datele!

O altă metodă neparametrică este cea a contigentei (propusă de Pearson) care permite stabilirea gradului de legătură între două variabile calitative ce nu pot fi cuantificate (numeric) sau dintre o variabilă cantitativă și una calitativă. Formula de calcul a coeficientului de contingență este :

Scanned with OKEN Scanner

$$c = \sqrt{\frac{\Phi^2}{1+\Phi^2}}$$

unde ϕ^2 este un factor dependent de frecvențele celor două variabile și de frecvențele combinațiilor dintre categorii și se determină după relația:

$$\Phi^2 = \sum_i \sum_j \frac{f_{ij}^2}{f_i f_j} - 1$$

Să exemplificăm (după Gugiuman Ana și colab, p.159-160) printr-o testare ce urmărește să stabilească dacă există o legătură între calificativele obținute de studenții de la educație fizică la practica de specialitate și notele de la examenul de Metodica educației fizice. Cum prima variabilă nu poate fi cuantificată, vom apela la aprecierea ei calitativă, în categoriile: slab, mediocru, bun și foarte bun. Aceleași calități le atribuim și notelor: slab (până la 5,50), mediocru (între 5,50 și 7,00), bun (între 7,00 și 8,50) și foarte bun (peste 8,50). Menționăm că metoda este valabilă pentru oricâte categorii atribuite variabilelor, indiferent dacă numărul acestora este egal sau inegal pentru cele două variabile.

Scanned with OKEN Scanner

Etapa următoare constă în alcătuirea tabelului de frecvențe:

	2	1	-	-	3
	3	7	3	-	13
	1	2	6	1	10
	-	1	2	6	9
	6	11	11	7	35

Pe ultima coloană apar frecvențele cumulate ale variabilei evaluare practică (f_i), iar pe ultima linie frecvențele cumulate pentru evaluarea teoretică f_j (6 pentru slab, 11 pentru mediocru, 11 pentru bun și 7 pentru foarte bun). Combinațiile acestor frecvențe (f_{ij}) apar la intersecțiile liniilor cu coloanele tabelului (f_{11} = 2, frecvența corespunzătoare unei evaluări practice slabe și teoretice tot slabe; f_{21} = 3, respectiv practic - mediocru și teoretic slab, ș.a.m.d). Astfel, cu aceste date, avem:

$$\Phi^2 = \left(\frac{2^2}{6}+\frac{1^2}{11}\right):3+\left(\frac{3^2}{6}+\frac{7^2}{11}+\frac{3^2}{11}\right):13+\left(\frac{1^2}{6}+\frac{2^2}{11}+\frac{6^2}{11}+\frac{1^2}{7}\right):10+$$

$$+\left(\frac{1^2}{11}+\frac{2^2}{11}+\frac{6^2}{7}\right):9-1=0,78$$

valoarea contingenței va fi

$$c = \sqrt{\frac{0,78}{1+0,78}} = 0,66$$

Valoarea obținută de 0,66 ne arată că există o legătură semnificativă între cele două evaluări.

Semnificația statistică a unui coeficient de corelație (ce ne informează despre gradul și sensul legăturii dintre variabile) exprimă riscul de eroare, mai exact șansa ca indicele respectiv să se obțină exclusiv datorită hazardului. Valorile coeficientului r se pot aprecia și intuitiv direct:

- valorile între 0 și ±0,20 denotă o corelație indiferentă între variabile;

- valorile între ±0,20 și ±0,40 denotă o corelație scăzută;

- între ±0,40 și ±0,70 se vorbește de cele mai multe ori de o corelație semnificativă (remarcabilă);

- în intervalul ±0,70 și ±1 corelația este ridicată, foarte puternică și semnificativă.

Începând de la corelațiile semnificative, dacă acestea sunt pozitive, atunci legătura dintre fenomene este directă

Scanned with OKEN Scanner

(creșterea unei variabile implică o creștere a celeilalte), iar dacă valoarea este negativă relația dintre fenomene este inversă (dacă o variabilă crește, cealaltă descrește).

Referitor la natura legăturii, în cazul unui coeficient de corelație semnificativ, aceasta rămâne la aprecierea cercetătorului, în funcție de variabile și de demersul științific în studiu.

În domeniul educației fizice și sportului, unde se prelucrează date ale organismului uman, deci valori determinate bio-psiho-social, supuse unui mare număr de factori întâmplători, se consideră ca semnificativi coeficienții pozitivi de la +0,70 în sus, spre +1.

Factori care influențează coeficienții de corelație:

- eșantionul ales să fie reprezentativ pentru populația de referință;
- eșantionul să respecte condițiile de omogenitate (vârstă, sex, pregătire), eterogenitate și de volum (număr cât mai mare de subiecți); după Garett (citat de Gugiuman Ana și colab., p.162) coeficientului r=0 îl corespunde o eroare probabilă de 0,15 la un colectiv format din 20 de subiecți, de 0,09 la un colectiv de 50 subiecți, de 0,06 la 100 subiecți și de 0,04 la 200 subiecți.

Scanned with OKEN Scanner

Stabilirea faptului existenței unei legături între două fenomene nu implică în mod necesar că între cele două fenomene există relații de cauzalitate. Există și corelații aparente, fără fundament real. G.U.Yule (cit. Porojan D., p.97) a stabilit că, între 1929 și 1937, în Marea Britanie valoarea coeficientului de corelație dintre numărul de receptoare radio și numărul deficienților mintali a fost de +0,988; este evident că, lăsând deoparte aspectul amuzant al determinării în cauză și apelând la bunul simț, nimic nu dă câștig de cauză existenței unei corelații reale între cele două variabile. Metodele statistice propun soluții și pentru aceste situații, prin intermediul corelațiilor parțiale sau de prim ordin (George A. F., p.389), ce au calitatea de a estima legătura dintre două caracteristici prin eliminarea efectului celei de-a treia (cu care primele două sunt corelate). Formula de calcul este următoarea:

$$r_{12\ 3} = \frac{r_{12} - r_{13} \times r_{23}}{\sqrt{1-r_{13}^2} \times \sqrt{1-r_{23}^2}}$$

unde $r_{12\text{-}3}$ este corelația dintre primele două variabile cu excluderea celei de-a treia, r_{12} este corelația dintre primele două variabile, r_{13} și r_{23} reprezintă coeficienții de

corelație dintre prima și a treia, respectiv dintre a doua și a treia variabilă.

Să considerăm două teste (X de inteligență și Y de calități motrice) și să le administrăm la un grup de elevi. Vom vedea că între cele două există o corelație semnificativă și că, la un grup de elevi de clasa a IV-a nivelul inteligenței și cel al calităților motrice va fi mult mai crescut decât la elevii de clasa I, de exemplu. Putem concluziona că, o corelație a celor două șiruri se datorează corelației cu o a treia, în cazul nostru vârsta. În acest caz vom folosi calculul corelației parțiale. Exemplificăm cu performanțele obținute la testarea X (inteligența), Y (calitățile motrice) și Z (vârsta), având r_{XY}=0,55 (sau r_{12}), r_{XZ}=0,60(sau r_{13}) și r_{YZ}=0,50(sau r_{23}):

$$r_{12-3} = \frac{0,55 - 0,60 \times 0,50}{\sqrt{1-0,60^2} \times \sqrt{1-0,50^2}} = +0,36$$

Aceasta înseamnă că indicele de corelație dintre primele două variabile (X și Y) prin eliminarea celei de-a treia (Z - vârsta) ar fi de 0,36; acest coeficient de mărime indică o corelație scăzută între cele două variabile. Considerând doar valoarea de 0,36 a lui r nu putem caracteriza destul

de precis relațiile dintre cele două variabile. Pentru aceasta, să abordă pe scurt coeficientul de determinare (r^2).

VII.1.2 Coeficientul de determinare (r^2)

Acest coeficient este un alt mod de exprimare a relațiilor dintre două variabile, în unele cazuri mai potrivit decât coeficientul de corelație Dacă, de exemplu, r=0,80 și r^2=+0,64, atunci 64% (r^2x100) din variația variabilei dependente (variabila efect) a fost explicată prin variația variabilei independente. Este mai ușor să se înțeleagă semnificația lui r^2 decât cea a lui r. De remarcat este relația dintre r și r^2: pe măsură ce r descrește față de valoarea sa maximă +1, r^2 descrește și mai repede, coeficientul de corelație fiind 0,70 atunci când cel de determinare este la jumătate:

0,81
0,64
0,49
0,36
0,25
0,16

Scanned with OKEN Scanner

Revenind la corelația parțială, avem pentru r_{12} (același cu $r_{1,1}$), $r_{12}^2=(0,55)^2=0,303$ și pentru $r_{12\text{-}3}^2 = (0.36)^2=0,127$. Diferența rezultată ținând, cont de vârstă, este 0,303-0,127=0,176. Dacă 0,303 reprezintă coeficientul de determinare r_{12}^2 dintre X și Y (deci 100%) atunci 0,176 (diferența) va reprezenta (0,176/0,303)x100=58%. Putem spune că o corelatie inteligență-calități motrice se datorează în 58 la sută din cazuri vârstei, diferența de 42% fiind asociată cu alți factori.

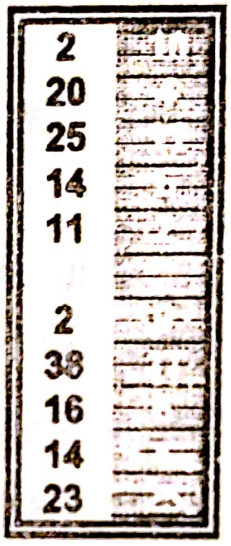
Exerciții

1. Să se calculeze coeficientul de corelație r și de determinare r^2 pentru următoarele date:

22
8
19
32
13
24
22
35
18
13

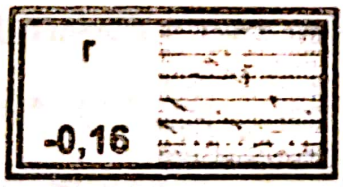
2. Să se calculeze coeficientul de corelație r și de determinare r^2 și pentru următoarele date:

3.

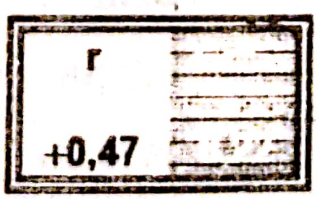


Răspunsuri:

1.



2.



Scanned with OKEN Scanner

CAPITOLUL VIII

TESTE DE SEMNIFICAȚIE

VIII.1. Ipoteze statistice

Cercetătorul care analizează diferite fenomene cu ajutorul metodelor statistice este interesat în primul rând de caracteristicile populațiilor din care provin eșantioanele de lucru Alcătuite după anumite reguli, eșantioanele nu au alt rol decât acela de a servi ca punct de plecare pentru o serie de inferențe asupra caracteristicilor (necunoscute) populațiilor, pornind de la caracteristicile cunoscute, dar fără relevanță intrinsecă, ale eșantioanelor. Aceste inferențe sunt numite ipoteze statistice. Ca orice ipoteză de lucru, și cele statistice trebuiesc verificate, iar verificarea lor se realizează prin intermediul unor procedee specifice numite teste statistice. Ipotezele statistice prezintă câteva particularități impuse de fundamentele probabilistice ale statisticii inferențiale inductive:

a) ipotezele statistice, în cea mai mare parte a lor, se formulează în termenii existenței unei diferențe între parametrii de sondaj și cei ai populației, a două sau mai multe populatii, între o distribuție de frecvență empirică și una teoretică sau între două distribuții empirice, verificarea ipotezelor înseamnă, în acest caz, evaluarea gradului de siguranță statistică (gradului de semnificație) a acestor diferențe. Concordanța diferitelor ipoteze cu realitatea se realizează în limitele unui anumit risc (probabilitate) de eroare acceptat (și stabilit) dinainte; acest risc este cunoscut sub denumirea de prag sau nivel de semnificație al testului folosit pentru verificarea ipotezelor statistice;

b) lipsa unei siguranțe depline (de 100%) conduce, implicit, la apriția erorilor în verificarea unei ipoteze statistice.

Aceste erori sînt de două tipuri:

a) respingerea ipotezei aunci când, în realitate, ea este adevărată (sau eroare de ordinul I);

b) acceptarea ipotezei atunci când ea este, în realitate, falsă (eroare de ordinul II, mai periculoasă decât precedenta, deoarece

Scanned with OKEN Scanner

conduce la generalizări false, erorile propagându-se ulterior în lanț).

În cadrul cercetărilor ameliorative, organizate în scopul perfecționării procesului instructiv educativ, cercetări care în mod curent își propun drept obiectiv studiul unor metode, mijloace, tehnici și procedee de lucru în cadrul lecțiilor sau metodelor de învățare, este necesar să se verifice validitatea ipotezei cerectării, respectiv măsura în care schimbarea produsă conduce sau nu la progresul real al elevilor sau sportivilor. Acest aspect impune compararea datelor obținute prin diverse metode de investigare specifice (probe de cunoștințe, de control, anchte, etc.) înainte și după încheierea experimentului (în cadrul aceluiași colectiv de subiecți sau între unul martor și celălalt experimental). Dacă diferențele observate se datorează doar întâmplării atunci crește riscul unei erori de gradul II; pentru a preveni această situație, în practica verificării ipotezelor statistice se oprează cu ipoteza care neagă existența acelei/acelor diferențe observate de cercetător. Această ipoteză inversă este cunoscută în statistică sub denumirea de ipoteza nulă (sau a diferenței nule) notată H_0. În raport cu aceasta, ipoteza de lucru devine o ipoteză alternativă (H_1) care va fi acceptată doar

Scanned with OKEN Scanner

dacă, în urma testelor specifice, ipoteza nulă este respinsă. Testarea ipotezei nule în locul celei alternative are ca principal avantaj micșorarea, în ceea ce privește ipoteza de lucru, a riscului erorii de gradul II.

VIII.2.

Parametrii care caracterizează o populație sînt niște valori fixe și sînt, de regulă, necunoscuți. De exemplu, vârsta medie a populației studenților de la facultățile cu profil de educație fizică și sport din întreaga țară în anul universitar 1995-1996 este un parametru cu o valoare anume, stabilă în intervalul temporal de definiție a populației. Această valoare poate fi cunoscută printr-un efort de inventariere a tuturor studenților dar ea poate fi și estimată cu ajutorul unor eșantioane extrase din populația respectivă. Prin comparație cu parametrii populației, cei de sondaj, variază de la un eșantion la altul extras din una și aceeași populație. Revenind la exemplul anterior este de așteptat ca pentru 10 eșantioane aleatoare prelevate din populația de studenti să obținem 10 vârste medii diferite, chiar dacă foarte apropiate între ele. Parametrii de sondaj

Scanned with OKEN Scanner

sunt mărimi cunoscute (calculabile) și acest fapt, combinat cu dificultatea (chiar imposibilitatea, uneori) cunoașterii parametrilor populației ("adevărați"), evidențiază importanța cunoașterii legăturii dintre cele două tipuri de parametrii. Cu alte cuvinte, este important să știm cât de reprezentativ este un eșantion pentru populația din care este extras, sau cu ce grad de precizie reușește un parametru de sondaj să estimeze (aproximeze) corespondentul său "adevărat". Cu cât eșantionul este mai mare, cu atât estimarea va fi mai precisă; dar, uneori, nu putem îndeplini această condiție deoarece în domeniul nostru se lucrează, în general, cu grupe mici, de 10-30 subiecți.

Eroarea standard a mediei unei colectivități este dependentă de volumul eșantionului (numărul de subiecți - n) și de abaterea standard (S) a distribuției și se calculează după relația:

$$\sigma_M = \frac{S}{\sqrt{n}}$$

Scanned with OKEN Scanner

Să exemplificăm, pentru a înțelege mai bine, cu un eșantion de n = 95 subiecți, abaterea standard S =±3,62 și media aritmetică $\bar{x}$=21,1:

$$\sigma_M = \frac{3,62}{\sqrt{95}} = 0,37$$

și să ne amintim de curba distribuției normale. Revenind la tabelul 5.1. (vezi capitolul V) căutăm în prima coloană (notată z) valoarea 1,9 și mergând în continuare spre dreapta până în dreptul valorii ,06 vom găsi valoarea 4750, ceea ce înseamnă că într-o distribuție normală 2x47,50%, adică 95% din rezultate, se încadrează între ±1,96σ; procedând identic, pentru valoarea de ±2,58σ vom afla că în acest interval vor fi cuprinse 99% din valori. Aceste intervale reprezintă limitele (intervalul de încredere) între care se află valoarea reală a mediei unei distribuții pentru populația de referință. Mai explicit, media unui eșantion se află între niște limite (să ne amintim că dacă prelevăm 10 eșantioane dintr-o populație, vom avea 10 medii foarte puțin diferite dar care se încadrează toate într-un același interval de variație) pe care le putem stabili cu o exactitate de 95%, respectiv 99% (dacă dorim să fim

Scanned with OKEN Scanner

mai precişi), dar niciodată de 100%. În domeniul nostru de activitate, pragul de semnificaţie cel mai îngăduitor pe care ni-l putem permite este de 95% (0,05). Astfel, în raport cu aproximaţia ei într-un eşantion, media se încadrează în intervalele:

$$\left(\bar{x} - 1{,}96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1{,}96\frac{\sigma}{\sqrt{n}}\right)$$

cu o încredere de 95% (sau probabilitate de 0,95, ceea ce înseamnă că putem greşi în 5% din cazuri)

şi

$$\left(\bar{x} - 2{,}5\frac{\sigma}{\sqrt{n}}, \bar{x} + 2{,}58\frac{\sigma}{\sqrt{n}}\right)$$

cu şansa de a greşi la un caz din 100 (99% încredere).

În exemplul nostru, σ_M=0,37 ceea ce înseamnă că media "adevărată" a populaţiei din care a fost extras eşantionul nostru de 95 de valori este:

$$21{,}1 \pm 1{,}96 \times 0{,}37$$

respectiv intervalul 20,37 ÷ 21,82

la pragul de semnificaţie 0,05 (5%)

şi

$$21{,}1 \pm 2{,}58 \times 0{,}37$$

respectiv intervalul 20,14 ÷ 22,05

Scanned with OKEN Scanner

la pragul de semnificație 0,01 (1%).

Trebuie să remarcăm faptul că ar fi corect ca în locul abaterii standard a eșantionului (S) să o folosim pe cea a populației (σ), care este necunoscută; pentru aceasta se aplică, de către cei care vor să fie foarte precișii în calculele lor, corecții ale lui S pentru a aproxima cât mai exact pe σ (corecția lui Bessel).

VIII.3. Comparația mediilor (testul lui Student - t)

Mediile a două șiruri de date pot să difere nesemnificativ (întâmplător) sau semnificativ; în cazul unei diferențe semnificative se poate presupune (și argumenta) prezența unui factor sistematic care a produs diferența, dar nu se poate demonstra că acest factor va produce același efect și la alte eșantioane, aparținând aceleiași populații statistice.

În practica experimentală se întâlnesc mai multe situații (în care se cere diferențierea dintre medii), cum ar fi: când șirurile sînt corelate sau necorelate, când abaterile

standard sînt cunoscute sau nu, etc., fiecare situație având propria formulă și algoritm după care se calculează parametrul "t". Parametrul t este cu atât mai mare cu cât diferența dintre medii este mai mare și cu cât abaterile standard ale șirurilor respective sînt mai mici. Când t calculat este mai mare decât o valoare t teoretică tabelată (vezi tabelul 8.1), în funcție de numărul de cazuri și pragul de semnificație despre care am vorbit anterior, la estimare, atunci ipoteza de nul (a diferenței întâmplătoare dintre medii) se infirmă; rezultă, cu o probabilitate acceptabilă, că mediile celor două eșantioane diferă semnificativ. Cea mai des întâlnită situație în practică este cea a metodei experimentale unde, prin testul lui Student, se urmărește argumentarea diferenței semnificative dintre media unui eșantion experimental față de media unui eșantion martor (de control).

VIII.3.1 Semnificația diferenței dintre medii pentru eșantioane distincte

Să presupunem că am testat două colective diferite (n_1, respectiv n_2 subiecți), provenite din aceeași populație,

cu aceeași probă de contol și le-am calculat mediile aritmetice ($\bar{x}$ și $\bar{y}$) și abaterile standard (S_x și S_y). Afirmăm, utilizând ipoteza nulă, că între cele două eșantioane nu există nici o diferență. Aplicăm testul lui Student pe baza relației:

$$t = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{S_x^2}{n_1} + \frac{S_y^2}{n_2}}} .$$

Să exemplificăm (după Gugiuman Ana și colab., p.168) cu două colective de 30, respectiv 36 de elevi care, la sfârșitul anului școlar au avut mediile generale pe clasă de 8,20 prima și 7,50 a doua, cu abaterile standard S_x=1,77 și S_y=1,86. Plecând de la ipoteza nulă, între mediile generale ale celor două clase de elevi nu există diferențe semnificative (sau altfel spus diferența este zero) ci doar factori întâmplători au determinat o oarecare diferență (de 0,70). Aplicând testul lui Student avem:

$$t = \frac{|8{,}20 - 7{,}50|}{\sqrt{\frac{1{,}77^2}{30} + \frac{1{,}86^2}{36}}} = \frac{0{,}7}{\sqrt{\frac{3{,}13}{30} + \frac{3{,}45}{36}}} = 1{,}79 .$$

Scanned with OKEN Scanner

În tabelul cu valorile repartiției Student vom căuta în prima coloană din stânga valoarea 64 - în tabelul nostru vom citi la 60 - (obținută din suma gradelor de libertate corespunzătoare fiecărui șir, respectiv 30-1+36-1) și vom găsi în dreptul valorii de probabilitate 0,05 cifra 2,00 , iar pentru o probabilitate de 0,99 (0,01) valoarea 2,66. Comparând valoarea calculată de noi t=1,79 cu cele teoretice, vom remarca faptul că în ambele cazuri aceasta este mai mică 1,79<2<2,66 ceea ce ne dă posibilitatea să acceptăm ipoteza nulă și să afirmăm că între elevii celor două clase nu este o diferență de valoare considerabilă, cu riscul de a greși în 5 respectiv 1% din cazuri.

VIII.3.2. Semnificația diferenței dintre medii pentru eșantioane corelate

De multe ori, un același grup de subiecți este supus la probe diferite (de exemplu, dacă se modifică performanțele unor sportivi la administrarea unui medicament) rezultând perechi de valori pentru fiecare

Scanned with OKEN Scanner

subiect, ce pot fi corelate. Dacă notăm cu D diferența dintre rezultatele obținute de fiecare individ din grup la cele două teste, testul t se calculează după formula:

$$t = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - \left(\sum D\right)^2}{n-1}}}.$$

Să considerăm că performanțele a zece sportivi au fost înregistrate înainte (X) și după (Y) un cantonament și s-au obținut următoarele rezultate:

Subiecți	X	Y
1	5	4
2	15	36
3	7	9
4	11	16
5	4	4
6	7	16
7	8	1
8	10	25
9	6	0
10	16	16
Total	89	127

Apelând la testul Student avem:

$$t = \frac{-13}{\sqrt{\frac{10 \times 127 - (-13)^2}{10-1}}} = \frac{-13}{\sqrt{\frac{1270-169}{9}}} = 1{,}175.$$

Scanned with OKEN Scanner

Valorile repartiției Student în funcție de probabilitate (p) și numărul gradelor de libertate (f)

Tabel 8.1

1	12,706	31,821	63,657
2	4.303	6,965	9,925
3	3,182	4,541	5,841
4	2,776	3,747	4,604
6	2,447	3,143	3,707
7	2,365	2,998	3,499
8	2,306	2,896	3.355
9	2,262	2,821	3.250
11	2,201	2,718	3,106
12	2,179	2,681	3,055
13	2,160	2,650	3,102
14	2,145	2,624	2,977
16	2,120	2,583	2,921
17	2,110	2,567	2,898
18	2,101	2,552	2,878
19	2,093	2,539	2,861
21	2,08	2,518	2,831
22	2,074	2,508	2,819
23	2,069	2,500	2,807
24	2,064	2,492	2,797
26	2,056	2,479	2,779
27	2,052	2,473	2,771
28	2,048	2,467	2,763
29	2,045	2,462	2,756
35	2,030	2,438	2,724
40	2,021	2,423	2,704
45	2,014	2,412	2,690
60	2,000	2,390	2,660
70	1,994	2,381	2,648
80	1,970	2,374	2,639
90	1,987	2,368	2,632
500	1,965	2,334	2,586
1000	1,962	2,330	2,581

Scanned with OKEN Scanner

De această dată căutăm în tabelul 8.1 în dreptul valorii obținute din numărul gradelor de libertate ale grupului nostru (nu se ține cont de semnul minus), deci la n-1=10-1=9 și vom găsi pentru probabilitatea de 0,05 valoarea 2,262 și respectiv 3,250 pentru o probabilitate de 0,01. Și în acest caz ipoteza de nul se confirmă, între cele două testări neexistând diferențe semnificative (1,175<2,262<3,250) ceea ce ne-ar permite să afirmăm că în condițiile de cantonament rezultatele obținute în final nu sunt diferite de cele inițiale. cu probabilitatea de a avea dreptate în 99% din cazuri.

Exercițiu:

Aplicați testul t pentru calcularea diferenței dintre medii pentru următoarele șiruri corelate:

T_1
31
19
22
26
36
30
29
36
33
34

Răspuns: t = 2,077 pentru 9 grade de libertate < 2.262 - nesemnificativ.

Scanned with OKEN Scanner

CAPITOLUL IX

PREZENTAREA DATELOR STATISTICE

Rezultatele prelucrărilor statistice se prezintă cu ajutorul tabelelor statistice și se reprezintă cu ajutorul graficelor.

IX.1. Tabelele statistice

Tabelele statistice constituie principala formă de prezentare a rezultatelor prelucrării statistice. Importanța tabelelor statistice constă în faptul că ele ne ușurează compararea și înțelegerea fenomenelor prezentate. Cu ajutorul lor caracterizăm sugestiv nivelul, structura, dinamica și repartiția teritorială a fenomenelor sociale, economice, etc.

IX.1.1. Elementele tabelului statistic

Orice tabel statistic este format din următoarele elemente structurale de bază:

Scanned with OKEN Scanner

1) *Macheta tabelului*, constituită din

a) *scheletul tabelului*, format dintr-o rețea de linii orizontale și verticale care, prin întretăiere, formează rânduri, coloane și rubrici (din întretăierea rândurilor și coloanelor rezultă rubricile, în care înscriem datele numerice);

b) *titlurile tabelului*, cuprind titlul general al tabelului, titlurile rândurilor și ale coloanelor;

2) *Subiectul tabelului*, arată despre ce colectivitate este vorba în tabel. El este dat de titlul general al tabelului și de titlurile rândurilor.Titlul general al tabelului se așează în partea superioară a tabelului și cuprinde obiectul despre care se vorbește în tabel, spațiul și timpul la care se referă datele, iar titlurile rândurilor se trec la capetele acestora;

3) *Predicatul tabelului*, ne arată acțiunea subiectului și este format din denumirea caracteristicilor prin care este descrisă colectivitatea cercetată, precum și din cifrele care arată nivelul lor de dezvoltare. El figurează de obicei la capetele coloanelor;

Scanned with OKEN Scanner

4) Datele numerice, se înscriu în rubricile tabelului şi se prezintă sub forma indicatorilor absoluţi şi derivaţi (mărimi relative, medii şi indici);

5) Nota explicativă, se înscrie uneori sub tabel şi ne indică sursa datelor şi metodele folosite pentru obţinerea diferiţilor indicatori. Ea însoţeşte tabelul în cazurile în care sînt absolut utile anumite explicaţii.

IX.1.2. Tipuri de tabele statistice

După modul de prelucrare şi prezentare, tabelele statistice pot fi:

1. *Tabele simple*, ce prezintă colectivitatea nedespărţită în grupe titpice. Ele se folosesc în scopuri informative şi de raportare;
2. *Tabele pe grupe*, prezintă colectivitatea despărţită în grupe omogene pe baza unei singure caracteristici de grupare, indiferent dacă aceasta se exprimă prin cuvinte sau cifre(tabelul 9.1, în care caracteristica de grupare se exprimă prin cuvinte - admisi, respinsi);

Scanned with OKEN Scanner

3. *Tabele combinate*, prezintă colectivitatea despărțită în grupe omogene pe baza a cel puțin două caracteristici de grupare folosite simultan (tabelul 9.2).

Situația după probele eliminatorii

Tabel 9.1

[illegible]	[illegible] candidați
admis	70
respins	30
[illegible]	[illegible]

Tabelele combinate sînt mai bogate în conținut deoarece ne oferă mai multe informații asupra colectivității prezentate.

Situația după probele eliminatorii

Tabel 9.2

Calificativul obținut	Numărul candidaților		
admis	70		20
respins	30		18
[illegible]	[illegible]		

Scanned with OKEN Scanner

Iată câteva reguli mai importante pentru alcătuirea tabelelor statistice:

- orice tabel statistic trebuie să aibă un titlu scurt, clar și complet, din care să rezulte fenomenul prezentat, timpul și spațiul unde acesta se desfășoară;
- alegerea formei tabelului (simplu, pe grupe, combinat) nu se face la întâmplare, ci în funcție de conținutul fenomenului ce urmează a fi prezentat și de scopul urmărit;
- subiectul se înscrie în tabel de obicei la capetele rândurilor, iar predicatul la capetele coloanelor;
- rândurile și coloanele trebuie să fie dispuse, în tabel, logic;
- trecerea indicatorilor în tabel se face în ordinea importanței lor;
- pentru fiecare indicator trebuie indicată unitatea de măsură; dacă aceasta este comună pentru toți indicatorii poate fi menționată deasupra tabelului.

Scanned with OKEN Scanner

IX.2. Reprezentările grafice

Graficele sînt un mijloc de reprezentare expresivă a datelor statistice cu ajutorul liniilor, punctelor, figurilor geometrice, figurilor naturale și simbolice.

Ele capătă o răspândire tot mai largă în statistică datorită următoarelor avantaje:

- înlesnesc compararea și memorarea datelor, datorită caracterului sugestiv și intuitiv;
- evidențiază nivelul, structura și deplasările structurale ale fenomenelor studiate;
- caracterizează dinamica fenomenelor;
- constituie un mijloc important de popularizare a datelor.

IX.2.1. Elementele reprezentării grafice

Elementele principale ale unui grafic sînt:

1) *Titlul graficului*, care trebuie să fie scurt, clar, complet și să corespundă cu titlul tabelului ale cărui date numerice le reprezentăm grafic. Din titlu trebuie să

Scanned with OKEN Scanner

reiasă fenomenul reprezentat, timpul și spațiul de desfășurare a fenomenului respectiv;

2) *Scara de reprezentare*, se compune din *suportul scării* (o linie) dispusă vertical sau orizontal, care este divizată printr-un șir de puncte numerotate. Ea are o limită inferioară și una superioară, iar distanța dintre limitele extreme ale scării se numește *lungimea scării*; distanța dintre grupele de puncte se numește *intervalul graficului*, iar distanța dintre valorile numerice (ale punctelor) se numește *interval numeric*;

3) *Rețeaua graficului*. este formată din linii paralele dispuse vertical sau orizontal, din linii încrucișate sau cercuri concentrice, care servesc la construirea graficului, purtând denumirea de *rețea ordinară*;

4) *Explicarea semnelor convenționale*, se face fie cu ajutorul unor *inscripții în chenar*, fie cu ajutorul unei *legendei*. La inscripții în chenar, explicațiile referitoare la graficul întocmit figurează într-un spațiu special deiimitat, prezentând astfel avantajul că ușurează citirea și reținerea datelor din grafic. legenda este o explicație sistematică a semnelor folosit în grafic. Ea se înscrie de regulă în afara graficului, sub rețea sau în partea dreaptă a graficului;

Scanned with OKEN Scanner

5) *Figurile cu care facem reprezentarea*, constituie elementul de bază al graficului. În funcție de natura seriei ale cărei date numerice urmează a fi reprezentate, aceste figuri pot fi geometrice, naturale, simbolice sau hărți.

IX.2.2. Reprezentarea grafică prin diagrame (figuri geometrice)

Diagramele se folosesc fie pentru reprezentarea structurii colectivității, fie pentru reprezentarea unor indicatori de volum. Condiția de bază care trebuie permanent avută în vedere în cazul diagramelor de volum este ca aria figurii geometrice să fie direct proporțională cu mărimea indicatorului reprezentat.

Cele mai utilizate sînt *diagramele prin pătrate* și cele prin *cercuri* (în cazul indicatorilor de volum), *dreptunghiul și cercul de structură*, *histograma sau poligonul de frecvență (vezi capII.2)* (pentru structura colectivității), *diagramele prin coloane sau benzi* (în variante bi- și tridimensioanle). Dar, să le analizăm pe scurt, pe fiecare dintre ele.

Scanned with OKEN Scanner

Să considerăm că în urma unei testări A, trei subiecți au obținut următorul punctaj:

Tabel cu numărul de puncte obținute în urma testării A

Tabel 9.3

Subiecți	X	Y	Z
Puncte	64	81	100

Pentru reprezentarea folosind o *diagramă prin pătrate* se extrage rădăcina pătrată din numărul punctelor obținute de fiecare din cei trei subiecți:

X: $l = \sqrt{64} = 8$ puncte;

Y: $l = \sqrt{81} = 9$ puncte;

Z: $l = \sqrt{100} = 10$ puncte.

În funcție de spațiul disponibil se alege unitatea de lungime a scării de reprezentare. De exemplu, ea poate fi de 1cm și atunci vom avea trei pătrate cu laturile (l) de 8, 9, respectiv 10cm, sau să zicem de 2cm și atunci lungimile laturilor vor fi 4, 4,5 respectiv 5cm:

Scanned with OKEN Scanner

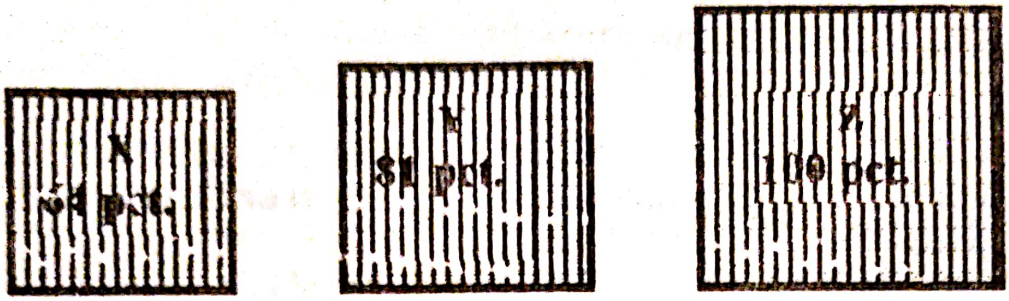


Fig. 9.1 Numărul de puncte obținute de cei trei subiecți în urma testului A

Pentru *diagrama prin cercuri* se parcurg aceleași etape cu deosebirea că pentru aflarea razei R a cercului de reprezentat se extrage rădăcina pătrată din raportul dintre valoarea obținută și π: (X) $R = \sqrt{\frac{64}{3,14}} = 4.514$ puncte, etc.

Reprezentarea grafică a structurii colectivității se face în funcție de seriile de distribuție, care sînt de două feluri: *calitative și cantitative*.

În cazul *seriilor cantitative* datele numerice se reprezintă grafic cu ajutorul histogramei și poligonului de frecvență, pe care le-am tratat la cap.II.

Reprezentarea structurii *seriilor calitative* se face cu ajutorul diagramelor structurale (*dreptunghi sau cerc de structură*); suprafețele acestora sînt egale cu 100% iar părțile lor sînt reprezentate prin porțiuni de suprafață.

Scanned with OKEN Scanner

Să considerăm următoarea serie calitativă (tabelul 9.4) rezultată din opțiunile făcute de studenții unei facultăți

Opțiunile exprimate de studenți pentru specializarea într-o disciplină sportiivă

Tabel 9 4

Fotbal	28,(8)
Handbal	20
Atletism	17,(7)
Volei	17,(7)
Baschet	15 (5)

de educație fizică și sport în ceea ce privește specializarea într-o disciplină sportivă

Construcția *dreptunghiului de structură* începe prin alegerea unei scări de frecvență dispusă vertical, pe care o divizăm în părți 10 egale, fiecare parte reprezentând

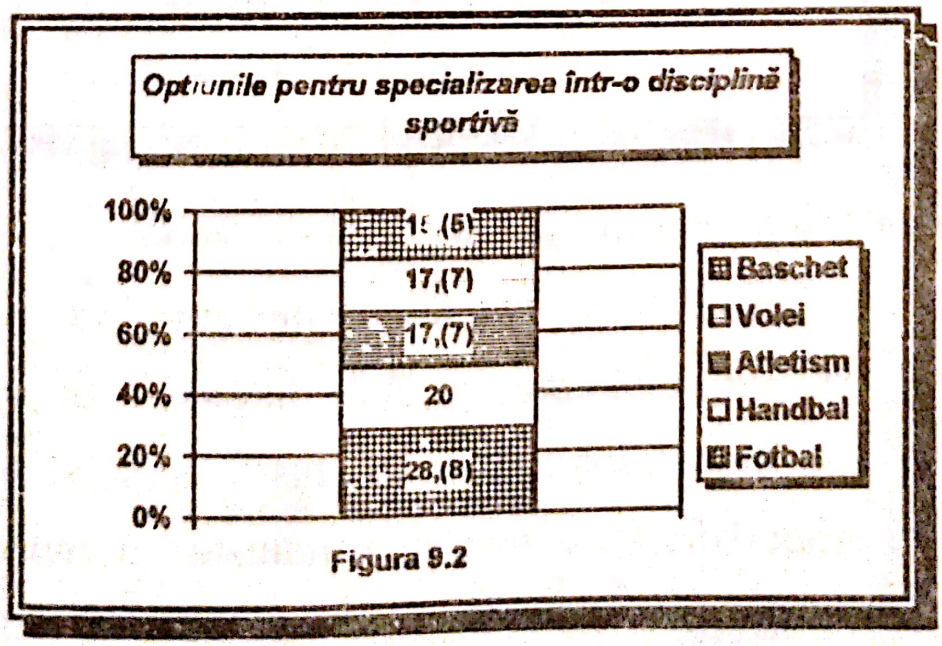


Figura 9.2

10% (fig.9.2).

Cercul de structură se obține prin reprezentarea ariei în procente volum total (100%). Părțile componente vor fi reprezentate prin sectoare de cerc ale căror arii vor fi proporționale cu mărimile relative de structură (360° reprezintă 100% din arie, iar sectoarele de cerc se obțin

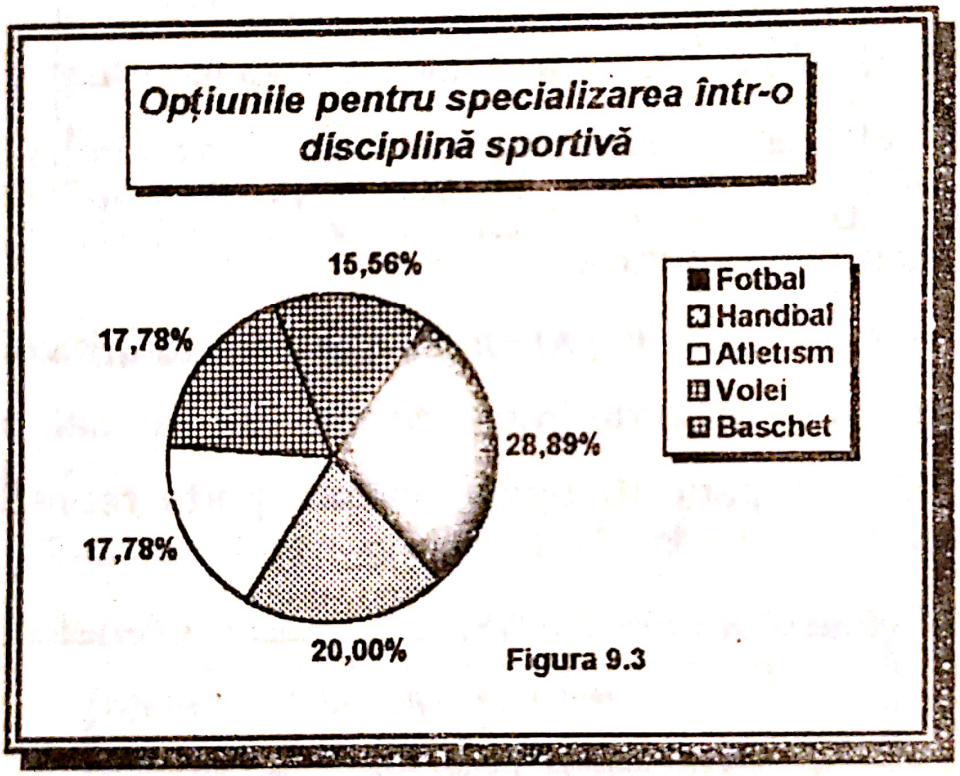


Figura 9.3

prin regula de trei simplă) (fig. 9.3).

Diagrama prin coloane este cea mai frecventă formă de reprezentare grafică întâlnită în statistică. Coloanele sînt dreptunghiuri cu baza egală și înălțimea direct proporțională cu mărimea indicatorilor reprezentați; ele sînt dispuse la distanță egală una de alta, distanță ce

Scanned with OKEN Scanner

reprezintă jumătate din lungimea bazei coloanelor (fig. 9.4).

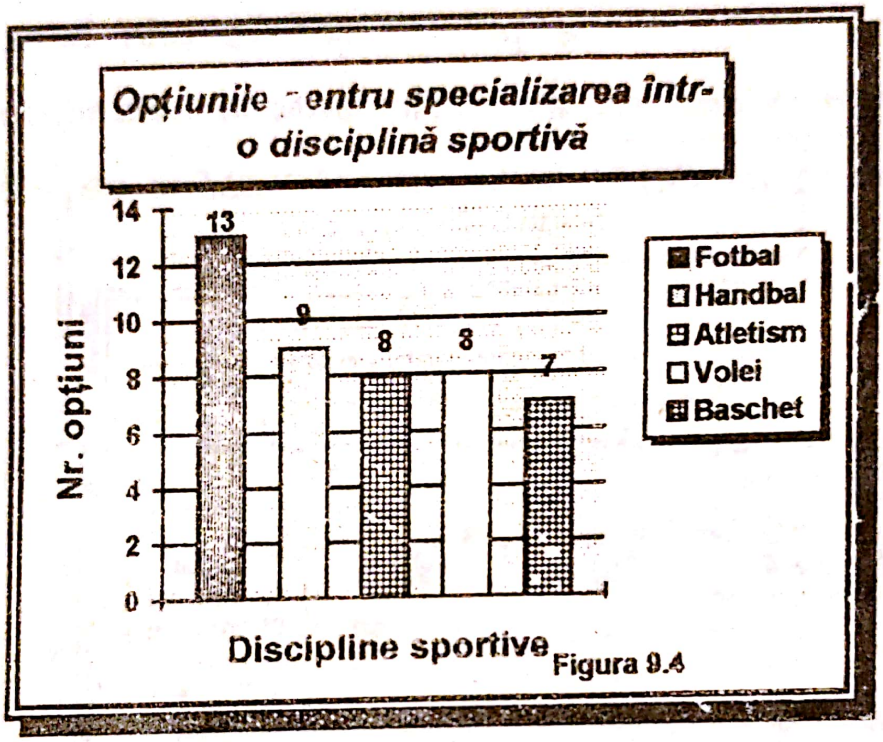


Figura 9.4

Diagrama prin benzi , spre deosebire de cea prin

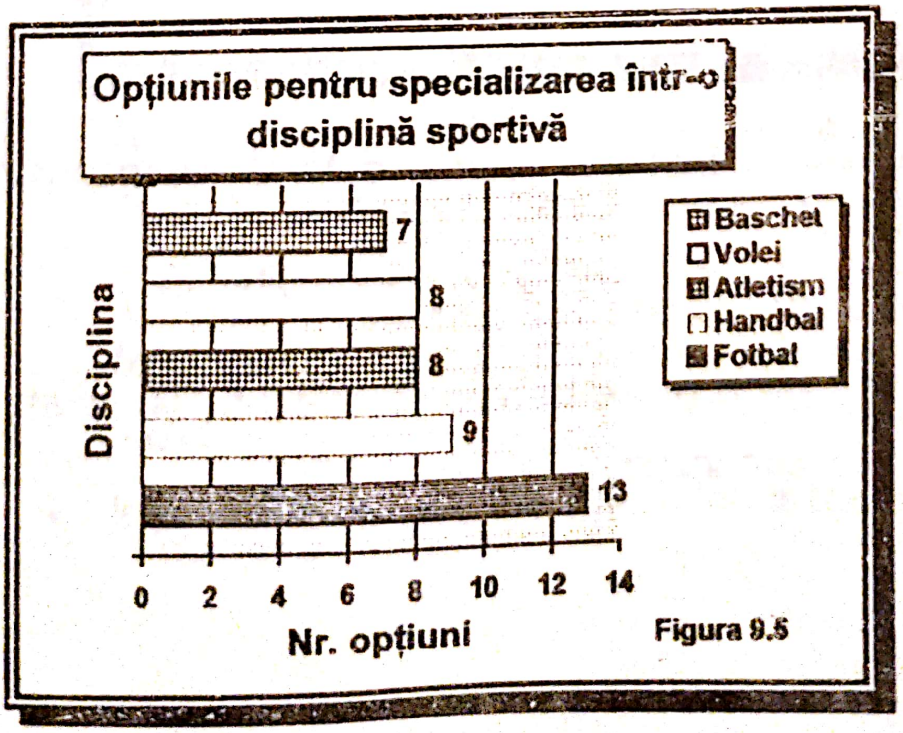
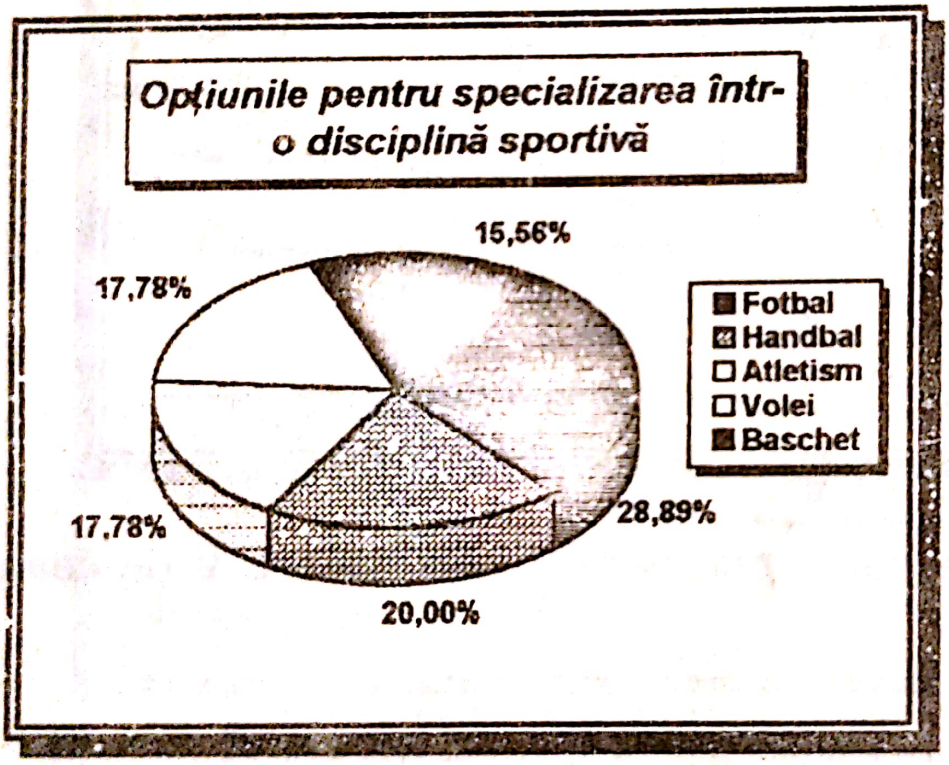


Figura 9.5

coloane, este realizată tot din dreptunghiuri, dar dispuse orizontal (fig. 9.5).

În reprezentare tridimensională, cercul de structură, diagrama prin coloane precum și cea prin benzi sînt aparent mai "profesionale", dificultatea constând doar (în cazul ultimelor două) în faptul că e destul de greoi să



citești valorile înscrise, tocmai datorită reprezentărilor în trei axe:

Reguli pentru construirea graficelor:

- orice grafic trebuie să aibă un titlu scurt, clar și cuprinzător;

Scanned with OKEN Scanner

- scara graficului trebuie să fie astfel construită încât să evidenţieze variaţia fenomenului reprezentat;

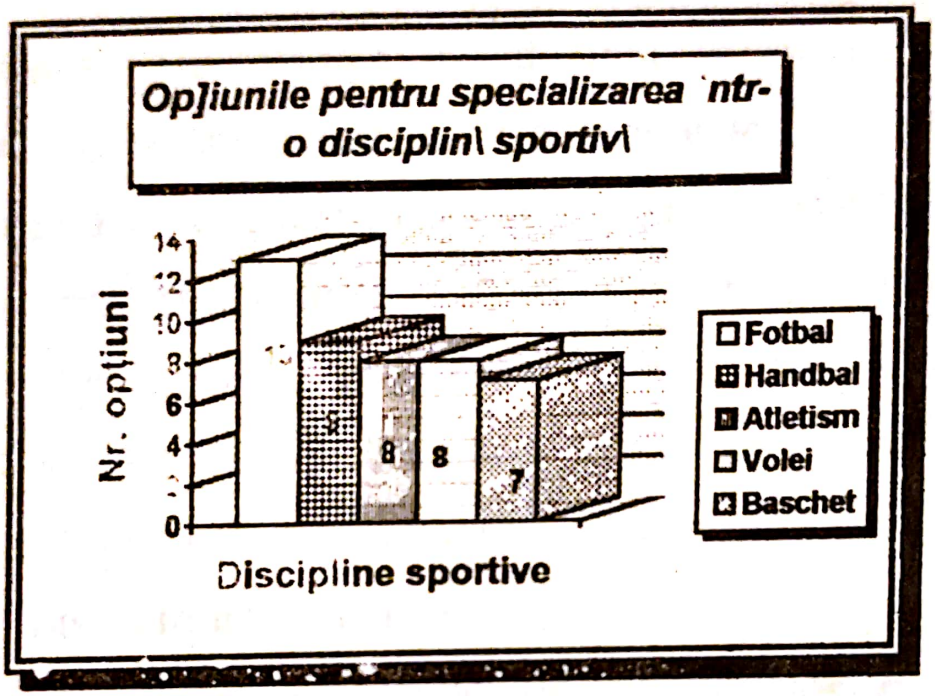


Diagrama tridimensională (prin coloane)

- notarea cifrelor pe scara verticală (axa Y) se face în partea stângă, iar pe scara orizontală (axa X) dedesupt. Scara trebuie să ne indice felul indicatorilor cu ajutorul cărora reprezentăm fenomenul: mărimi absolute, relative, medii, indici;
- legenda se aşează, de obicei, în partea dreaptă.

Scanned with OKEN Scanner

CAPITOLUL X

STATISTICA ŞI SISTEMELE DE CALCUL

Realizare importantă a industriei de tehnică de calcul din ultimii ani, calculatorul personal are utilizatori din cele mai diverse categorii sociale: ingineri, economişti, tehnologi, proiectanţi, matematicieni, fizicieni, chimişti, medici, meteorologi, profesori, funcţionari, artişti, muncitori, sportivi, cercetători, elevi, etc. Explicaţia constă în virtuţile acestui echipament, dintre care se detaşează:

- viteza mare de calcul;
- accelerarea procesului decizional folosind un mare volum de informaţii stocat şi prelucrat corespunzător;
- creşterea calităţii cercetării ştiinţifice şi proiectării tehnologice;
- folosirea cu rezultate superioare a resurselor materiale şi umane;
- perfecţionarea serviciilor şi, nu în ultimul rând,
- paleta largă de divertisment oferită de "jocurile pe calculator" cu o grafică şi ilustrare muzicală extrem de atractive.

Scanned with OKEN Scanner

Se alătură acestor calități și faptul că un calculator personal (PC) are un gabarit mic și o greutate redusă, este interactiv și ușor de deplasat și montat unde se dorește și se caracterizează printr-un preț accesibil.

În prezent, calculatorul nu mai este un obiect prețios care provoacă uimire, ci tinde mai curând să capete rolul de instrument util și foarte răspândit, devenind o componentă de neînlocuit în viața de toate zilele, de la locul de muncă și până la momentele de folosire a timpului liber. Și totuși, în intervalul de numai câțiva ani, el a modificat profund viața socială și a transformat orice tip de informație, cu o capacitate de penetrație pe care nici măcar inventatorii lui nu și-ar fi închipuit-o în urmă cu numai 50 de ani.

Computerul nu este totuși doar un "calculator", în accepțiunea literală a termenului, în sensul că el nu tratează numai numere, ci prelucrează, în accepțiunea cea mai largă a cuvântului, informații. În primul rând a cucerit foarte rapid rangul de produs comercial la prețuri din ce în ce mai accesibile. Calculatorul - deși este printre ultimii apăruți în familia produselor electronice - a devenit o realitate caracteristică a societăților industriale dezvoltate (sau postindustriale), astfel încât "putem evalua gradul de

Scanned with OKEN Scanner

dezvoltare a unei societăți după numărul de calculatoare utilizate"(Giovannini G., p.216). Capacitatea sa uluitoare de a prelucra informații, de a memora cantități imense de date la costuri tot mai scăzute și de interactivitate asigură calculatorului un rol fundamental în însăși evoluția societății de mâine. Generația actuală de calculatoare personale posedă programe (aplicații) special construite pentru a putea fi folosite de orice persoană și nu numai de către specialiști (cei pe care proaspeții inițiați și nu numai ei, îi privesc ca pe niște "guru" ai domeniului). Cu cât PC-uri-le sînt mai rapide și mai complexe, cu atât programele sunt mai ușor de utilizat și pot realiza mai multe operații.

Datorită computerelor, modalitatea statistică de prelucrare a datelor și chiar a informațiilor a devenit una din cele mai răspândite și utilizate, în toate domeniile abordabile științific. Și în domeniul educației fizice și sportului, comunitatea academică recunoaște și apreciază importanța instrumentului de cunoaștere statistic, interesul manifestându-se preponderent pentru statistica aplicată pe PC, respectiv soft-uri profesionale pentru prelucrari statistice de date.

Produs al arhicunoscutei, de acum, firme Microsoft, sistemul Windows se bucură în prezent de o mare

Scanned with OKEN Scanner

popularitate și în țara noastră, fiind cea mai folosită interfață grafică utilizator.

Între aplicațiile care rulează sub Windows și care fac parte din pachetul MS-Office, poate cea mai performantă este Excel (Fig.10.1) care oferă posibilitățl de lucru cu foi de calcul tabelar (*spreadsheet-uri*), alcătuind o agendă de lucru impresionantă, formată din 16 asemenea foi, constituite fiecare dintr-un număr consistent de linii (16384, numerotate 1, 2, 3, ... , 16384) și de coloane (256 notate cu A, B, C, AA, AB, ... , IV).

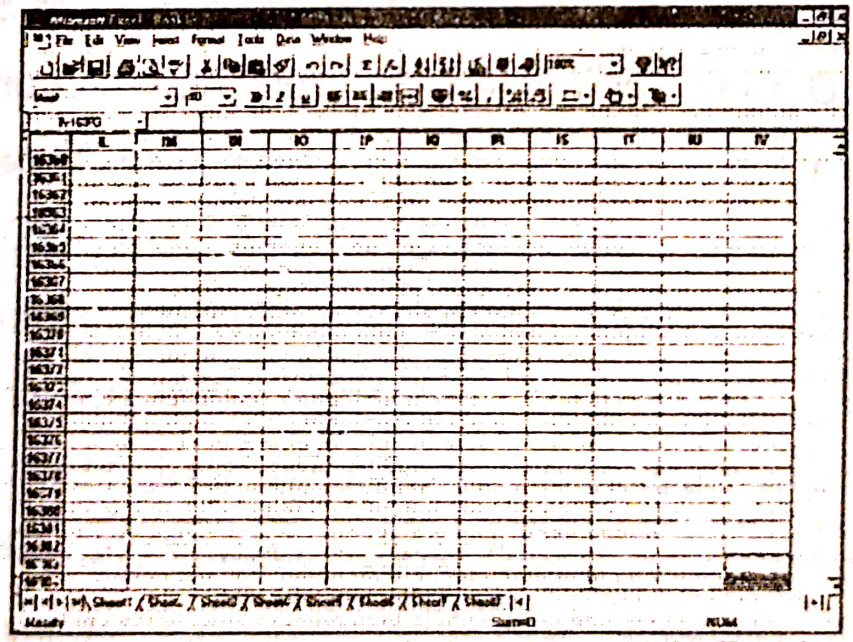


Fig. 10.1 Fereastra de lucru curentă în Excel (cu evidențierea ultimei celule)

Această aplicație oferă utilizatorilor domeniului nostru (fie ei studenți, cadre didactie sau cercetători consacrați) diverse facilități cum ar fi calcule și analize

Scanned with OKEN Scanner

statistico-matemetice, interpretări ale datelor obținute în urma acestor calcule, proiectarea bazelor de date și probleme de decizie și raportări.

Un dezavantaj major îl constituie, pentru specialiștii noștrii, lipsa unei documentații care să facă "lumină" asupra utilității acestui produs și ale posibilităților lui de folosire de către studenții sau cadrele didactice din facultățile de profil.

Funcții statistice predefinite de largă utilizare în aplicația Excel

Tabel 10.1

Denumirea funcției
1. Average
2. Median
3. Mode
4. Max
5. Min
6. Avedev
7. Devsq
8. Stdev
9. Quartile
10. Standardize
11. Correl și Pearson
12. Confidence
13. Ttest

Numărul funcțiilor statistice pe care le considerăm că sunt facile este de 13 funcții de largă utilizare, la care

se pot adauga permanent altele noi legate de necesitățile și cunoștințele fiecăruia (Tabel 10.1). Procentual, acesta reprezintă doar 18,3% din totalul de 71 de funcții statistico-matematice predefinite disponibile în această aplicație.

Așa cum spuneam anterior referitor la problemele de raportări și de decizie, XL se dovedește un instrument foarte puternic în ceea ce privește posibilitățile de exprimare grafică a diverselor rezultate. Numai dacă amintim faptul că oferă posibilitatea de a alege între 15 moduri distincte de reprezentare a graficelor, care la rândul lor, prin variante, conduc la o "paletă" ce sumează 102 posibilități bi- și tridimensionale de a te face înțeles și altfel decât prin cifre care, uneori, sunt mai relevante într-o astfel de vizualizare, ne dăm seama de reala utilitate a acestui produs soft.

Și aproape nimic din tot ceea ce am spus până aici nu ar avea valoare dacă nu am aminti despre facilitățile pe care le oferă această aplicație în crearea și gestionarea bazelor de date, care, până la urmă, reprezintă materialul de prelucrat și de reprezentat. O bază de date este un instrument folosit pentru depozitare, organizare și extragere de informații. Scopul creării unei baze de date este nu atât acela de a calcula noi valori, ci mai degrabă

Scanned with OKEN Scanner

de a stoca o mulțime de informații într-o manieră coerentă. Păstrarea unei mari cantități de date într-un format tabelar reprezintă doar o mică parte a "filosofiei" bazelor de date în programul Excel.

Flexibilitatea ordinii înregistrărilor este exact ceea ce se cere pentru a putea rezolva toate problemele legate de gestionarea datelor. Comanda Sort (din meniul Data) oferă această flexibilitate, acesteia alăturîndu-l-se AutoFilter pentru a filtra baza de date, cu scopul de a afișa numai acele înregistrări care conțin o anumită valoare.

În urma tratării tutror acestor aspecte, sintetic, se poate concluziona, în ceea ce privește aplicația Excel, că:

- este deosebit de utilă cunoașterea cât mai aprofundată a posibilităților pe care le oferă în utilizare domeniului nostru de activitate;
- realizează formarea utilizatorului (cea mai mare parte a acestora fiind în prezent studenții) în vederea creșterii valorii studiilor științifice;
- indirect, crește obiectivitatea studiilor științifice, prin exprimări cifrice (sau conexiuni logice) a datelor;
- crește precizia datorită facilităților de căutare și sortare disponibile pentru baze de date;

Scanned with OKEN Scanner

- ordonează întreaga activitate de lucru cu calculatorul, ceea ce duce implicit la scăderea sensibilă a timpului de lucru.

Foarte puțini din utilizatori știu însă că, o aplicație orientată în principal pe desenare de grafice, include și posibilități de lucru cu funcții, între acestea incluzându-se și funcțiile statistice. Aceasta se numește CorelCHART! (aplicație prezentă în pachetul CorelDRAW!, produs al firmei Corel Corporation) și este un program de grafică (desen) vectorială - "are o interfață intuitivă care îl face ușor de învățat dar, totodată, greu de stăpânit" (Gray D., p.1); el pune la dispoziția utilizatorului, printre altele,

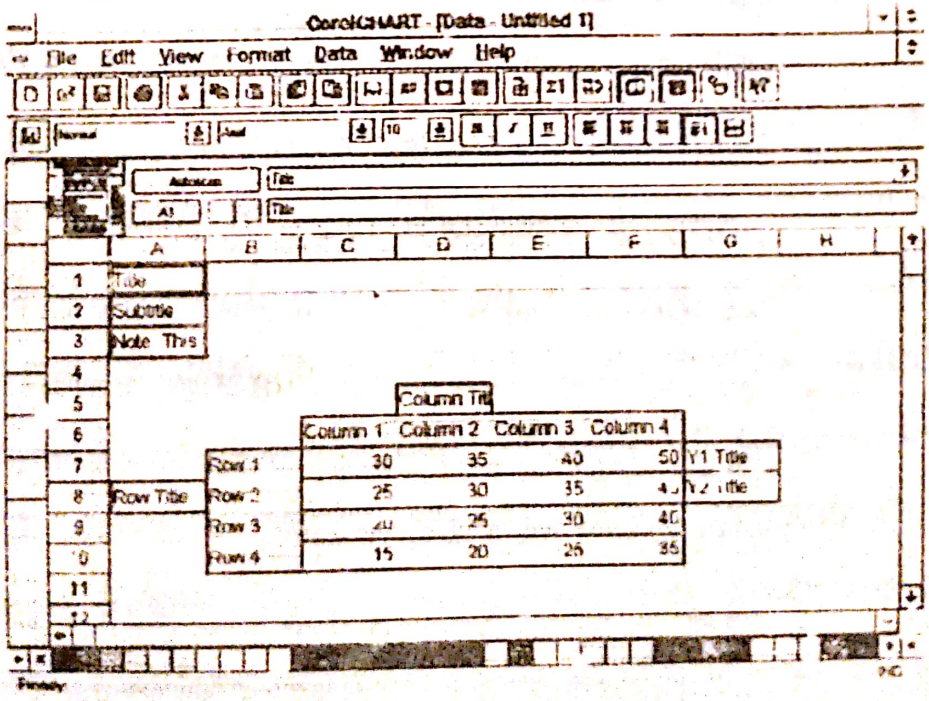


Fig.10.2

Scanned with OKEN Scanner

instrumente pentru scrierea formulelor şi funcţii pentru calcularea datelor din diagrame (fig.10.2).

CorelCHART! Conţine peste 300 de funcţii pentru următoarele tipuri de operaţii:

- funcţii de conversie;
- funcţii pentru oră şi dată;
- funcţii DDE (Dynamic Data Exchange) şi externe;
- funcţii pentru calcule inginereşti;
- funcţii pentru calcule financiare;
- funcţii logice;
- funcţii matematice şi trigonometrice;
- funcţii statistice, etc.

Funcţiile din această aplicaţie lucrează ca şi cele din aplicaţiile cele mai răspândite de calcul tabelar; ea pune la dispoziţie o foaie de calcul cu 256 de coloane (etichetate cu litere de la A la IV) şi 16384 de linii (notate cu cifre de la 1 la 16384). Utilizatorii cu o oarecare experienţă în lucrul cu aplicaţiile de calcul tabelar (Excel, Lotus 1-2-3) vor constata că formulele şi funcţiile din CorelCHART! le sunt foarte familiare!

Scanned with OKEN Scanner

Pentru a putea opera cu oricare din funcțiile prestabilite enunțate mai sus se apelează meniul Data, comanda Enter Formula (F_{12}) meniul derulant Functions;

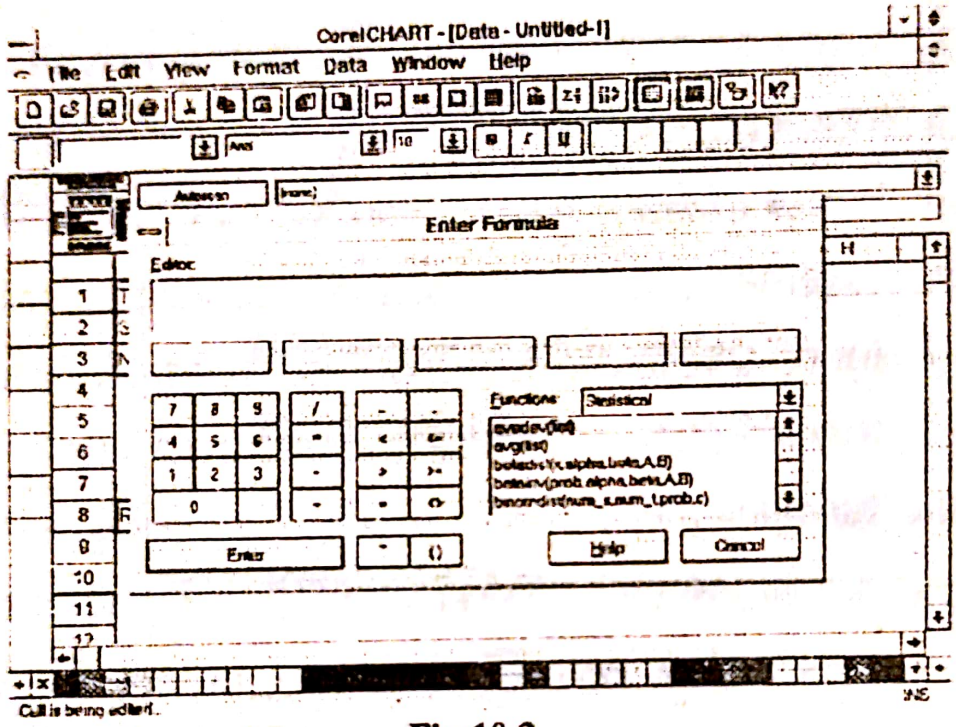


Fig.10.3

în cazul nostru ne vom opri la cele statistice (Statistical, fig.10.3).

Aplicația pune la dispoziție pentru calculele statistico-matematice a șirurilor de date recoltate în urma măsurilor un număr destul de mare de funcții (65), dar inferior celor din aplicația "rivală" Excel (care în varianta 5.0 cuprindea 71 de asemenea funcții, număr ce a crescut ulterior în varianta recentă '97). Evident, un cercetător din domeniul nostru de activitate care este în același timp și utilizator al acestei aplicații, nu va fi nevoit aproape

Scanned with OKEN Scanner

niciodată să le cunoască și să le folosească pe toate. Tocmai de aceea, pentru a veni cu un minim de informație în această direcție, vom trece în revistă, foarte pe scurt, pe câteva dintre acestea, pe cele pe care le socotim a fi mai des utilizate (tabel 10.2):

Funcții statistice în CorelCHART! De largă utilitate

Tabel 10.2

Denumirea funcției	Utilitate
1. avg	
2. median	
3. mode	
4. max	
5. min	
6. avedev	
7. devsq	
8. std	
9. quartile	
10. Standdize	
11. correl și pearson	
12. ttest	

Opinia noastră este că în aceast program, deși destinația principală este crearea de grafice și nu calculul tabelar, Help-ul este construit într-o engleză mai aceesibilă, mai concis și cu exemple mai clare. Cu alte cuvinte - dar nedorind sub nici o formă să îi criticăm pe cei

Scanned with OKEN Scanner

de la Microsoft - în acest caz help-ul ajută într-adevă utilizatorul!

Nu putem omite, în încheiere, numirea și a altor programe care înlesnesc munca cercetătorului, cum ar fi SPSS, Statistica for Windows, etc. Ar mai fi de precizat că, "din păcate, ceea ce, de cele mai multe ori, este trecut sub tăcere, este faptul că acest instrument de cunoaștere, dacă nu este folosit adecvat, duce la concluzii eronate sau abuzive. [...] Mai important decât a ști pe care butoane să apăsăm este, însă, să știm ce semnificație au cifrele seci ale statisticii și cât de mult ne putem baza pe ele (Gagea A., p.40-41)."

Scanned with OKEN Scanner

REFERINȚE BIBLIOGRAFICE

1. **Apetrei, M. și colab.** -*Elemente de statistică cu aplicații în geografie*, Editura Universității "Al.I.Cuza", Iași,1996;
2. **Clarke, H.H.** -*Application of measurement to health an physical education (fifth edition)*, Prentince-Hall, Inc., New Jersey, 1976;
3. **Dragnea, A.** -*Măsurarea și evaluarea în educație fizică și sport*, Editura Sport-Turism, București, 1984;
4. **Epuran, M.** -*Metodologia cercetării activităților corporale*, Editura A.N.E.F.S., București, 1992;
5. **Ferguson, G.A.** -*Statistical analysis in psychology and education(second edition)*, McGraw-Hill Book Company, New York, 1966 ;
6. **Gagea, A.** -*Curs de informatică și statistică*, Editura A.N.E.F.S., București, 1996;
7. **Gheorghiu, R.C.** -*Excel 5.0 (sub Windows) poate fi o aplicație prietenoasă și pentru studentii facultăților de educație fizică si sport*, Rev. Sport și Societate, Iași, 1/1996;
8. **Gheorghiu, R.C.** -*Importanta cunoașterii funcțiilor statistice din Excel 5.0 (sub Windows) cu aplicabilitate în lucrările științifice studențești din domeniul educației fizice și sportului*, Analele universității "Ovidius" Constanța, vol.I, anul I.1996;
9. **Giovannini, G. (coord.)** -*Istoria mijloacelor de comunicare în masă*, Ed. Tehnică, București, 1989;
10. **Gray, D.** -*Secretele programului CorelDRAW! 5*, Ed.Teora, Bucuresti, 1996;
11. **Gugiuman, Ana, (coord.)** -*Intoducere în cercetarea pedagogică*, Editura Tehnică, Chișinău, 1993;
12. **Guilford, J.P., Fructher B.** -*Fundamental statistics in psychology and education (fifth edition)*, McGraw-Hill Book Company, New York, 1973;
13. **Iosifescu, M. și colab.** -*Mică enciclopedie de statistică*, Editura Științifică și enciclopedică, București, 1985;
14. **Ivănescu, I. (coordonator)** *Statistică*, Editura Didactică și pedagogică, București, 1980;
15. **Jaba, Elisabeta** - *Statistica*, Editura Libris, Iași, 1996
16. **Mihoc, Gh. și colab.** *Teoria probabilităților și statistică matematică*, Editura Didactică și pedagogică, București, 1970;
17. **Peterson, R.** -*Excel pentru Windows 95*, Ed. Teora, București, 1996.
18. **Tudös Șt.** -*Elemente de statistică aplicată*, Ministerul Tineretului și Sportului, București, 1993

Scanned with OKEN Scanner

Cuprins

CAPITOLUL I

NOTIUNI GENERALE ____ 3

I.1. Scurt istoric ____ 3

I.2. Obiectul, metoda și etapele cercetării statistice ____ 7

I.3. Principalele noțiuni folosite în cercetarea statistică ____ 11

CAPITOLUL II

DISTRIBUȚIA DE FRECVENȚE

ȘI REPREZENTAREA EI GRAFICĂ ____ 14

II.1. Distribuția de frecvențe ____ 14

II.2. Reprezentarea grafică a distribuției ____ 22

Exerciții ____ 28

CAPITOLUL III

PARAMETRII TENDINȚEI CENTRALE ____ 30

III.1. Media aritmetică simplă și ponderată ____ 31

III.1.1. Media aritmetică simplă ____ 32

III.1.2.Media aritmetică ponderată ____ 33

III.1.3 Proprietățile mediei aritmetice ____ 36

III.1.4. Media aritmetică pentru mai multe șiruri ____ 39

III.2 Media armonică simplă și ponderată ____ 40

III.2.1. Media armonică simplă ____ 40

III.2.2. Media armonică ponderată ____ 41

III.3. Mediana și modulul ____ 42

III.3.1. Mediana(M_e) ____ 42

III.3.2 Modulul (M_o) ____ 46

Exerciții ____ 48

CAPITOLUL IV

PARAMETRII DISPERSIEI ____ 50

IV.1. Amplitudinea de variație(A) ____ 51

IV.2. Cuartilele și abaterile cuartilice ____ 52

Scanned with OKEN Scanner

IV.3. Abaterea medie(A_m) ____ 54

IV.4. Dispersia (variația, σ^2) ____ 55

IV.5. Abaterea standard (abaterea medie pătratică) ____ 58

Exerciții ____ 61

CAPITOLUL V

DISTRIBUȚIA NORMALĂ ____ 62

V.1.Distribuția (curba) normală și caracteristicile sale ____ 62

V.2. Parametrii formei ____ 69

V.2.1. Indici de asimetrie ____ 69

V.2.2 Indicii de aplatizare (de exces) ____ 71

Exerciții ____ 73

CAPITOLUL VI

VARIABILELE NORMATE (ETALONAREA) ____ 74

VI.1. Variabilele Z ____ 75

VI.2. Variabilele T ____ 78

VI.3. Variabilele H ____ 79

Exerciții ____ 80

CAPITOLUL VII

CORELAȚIA STATISTICĂ ____ 81

VII.1. Coeficientul de corelație și de determinare ____ 82

VII.1.1 Coeficientul de corelație ____ 82

VII.1.2 Coeficientul de determinare (r^2) ____ 99

Exerciții ____ 100

CAPITOLUL VIII

TESTE DE SEMNIFICAȚIE ____ 102

VIII.1. Ipoteze statistice ____ 102

VIII.2. Estimarea ____ 105

VIII.3. Comparația mediilor (testul lui Student - t) ____ 109

VIII.3.1 Semnificația diferenței dintre medii pentru eșantioane distincte ____ 110

Scanned with OKEN Scanner

VIII.3.2. Semnificaţia diferenţei dintre medii pentru eşantioane corelate 112
Exerciţiu 115

CAPITOLUL IX

PREZENTAREA DATELOR STATISTICE 116

IX.1. Tabelele statistice 116
IX.1.1. Elementele tabelului statistic 116
IX.1.2. Tipuri de tabele statistice 118

IX.2. Reprezentările grafice 121
IX.2.1. Elementele reprezentării grafice 121
IX.2.2. Reprezentarea grafică prin diagrame (figuri geometrice) 123

CAPITOLUL X

STATISTICA SI SISTEMELE DE CALCUL 131

REFERINŢE BIBLIOGRAFICE 143

Redactor:
MIHAELA POPOVICI

Format: 61×86/16
Coli tipo: 9,50
Apărut: 1998
Comanda: 92

Tiparul executat la Întreprinderea Poligrafică „A.I. Cuza"

Scanned with OKEN Scanner